

## **DATA MINING MEMPREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE K-NEAREST NEIGHBORS (KNN) STUDI KASUS UNIVERSITAS PGRI MAHADEWA INDONESIA**

**I Putu Yogista Putra Atmaja<sup>1</sup>, Gde Iwam Setiawan, SE., M.Kom<sup>2</sup>, I Wayan Dika, S.E., M.Pd<sup>3</sup>, Ida Ayu Putu Febri Imawati<sup>4</sup>**

<sup>1</sup>Universitas PGRI Mahadewa Indonesia, Denpasar, Indonesia; yogiputra0586@gmail.com

<sup>2</sup>Universitas PGRI Mahadewa Indonesia, Denpasar, Indonesia; iwansetiawan@mahadewa.ac.id

<sup>3</sup>Universitas PGRI Mahadewa Indonesia, Denpasar, Indonesia; dika@mahadewa.ac.id

<sup>4</sup>Universitas PGRI Mahadewa Indonesia, Denpasar, Indonesia; febri@mahadewa.ac.id

### **ARTICLE INFO**

#### **Article history**

Received Agustus 29, 2023

Revised September 11, 2023

Accepted October 04, 2023

Available online October 28, 2023

Keywords: Data Mining, K-Nearest Neighbors (KNN), Python

*Copyright © by Author. Published by  
Fakultas Teknik dan Informatika  
Universitas PGRI Mahadewa Indonesia*

**Abstract.** Graduation is a significant milestone in education, and it is a crucial assessment factor for ensuring higher education accreditation. The K-Nearest Neighbor (KNN) algorithm classifies objects based on learning data, with a minimum and maximum number of training datasets. The algorithm normalizes patterns, calculates Euclidean distance, votes from the smallest euclidean distance, and determines the classification results. The Student Graduation Prediction Model uses the KNN method to help assess students' graduation accuracy and accreditation.

### **PENDAHULUAN**

Kelulusan merupakan hal yang sangat membahagiakan bagi semua orang yang menempuh jenjang pendidikan. Kelulusan mahasiswa menjadi salah satu faktor penilaian untuk meningkatkan akreditasi perguruan tinggi. Untuk menjamin mutu dan kualitas lulusan dari perguruan tinggi dibutuhkan penilaian yang memenuhi suatu syarat yang baku atau kriteria tertentu. Hal tersebut diatur dalam Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 5 Tahun 2020 mengenai akreditasi sebuah lembaga pendidikan yang diwajibkan untuk melakukan proses akreditasi melalui Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) (KEMENDIKBUD, 2020). Akreditasi sendiri merupakan salah satu faktor penting mahasiswa dalam memilih suatu jurusan dan perguruan tinggi.

Data Mining (Penambangan data) merupakan bidang ilmu multidisiplin, menggambarkan area-area kerja yang termasuk didalamnya adalah teknologi basis data, pembelajaran mesin, statistik, pengenalan pola, pengambilan informasi, jaringan saraf tiruan, sistem berbasis pengetahuan, kecerdasan buatan, komputasi kinerja-tinggi, dan visualisasi data (Meilina, 2015). Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database Jurnal Ilmu Komputer Indonesia yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Data mining biasa juga disebut dengan "data atau *knowledge discovery*" atau

menemukan pola tersembunyi pada data. Data mining adalah proses dari menganalisa data dari prespektif yang berbeda dan menyimpulkannya ke dalam informasi yang berguna (Kadek et al., 2019). Data mining merupakan sebuah proses pengumpulan informasi dan data yang penting dalam jumlah yang besar atau *big data* untuk mencari informasi penting didalam data tersebut. Salah satu algoritma yang digunakan dalam data mining adalah K-Nearest Neighbors (KNN).

Penelitian ini penulis lakukan karena proses penentuan kelulusan mahasiswa masih secara manual, mengakibatkan lamanya proses penentuan itu berlangsung dan lama juga hasil yang di peroleh. Pada penelitian ini penulis berupaya melakukan mining data dengan menerapkan salah satu algoritma yaitu K-Nearest Neighbors. Metode KNN melakukan klasifikasi terhadap objek berdasarkan data latih (*training*) yang menggunakan jarak terdekat atau kemiripan terhadap objek tersebut. Pada fase pembelajaran, algoritma ini melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data pengujian (yang klasifikasinya tidak diketahui). Setelah jarak dari vektor yang baru terhadap seluruh vektor data pembelajaran dihitung dan sejumlah K buah yang paling dekat diambil, selanjutnya klasifikasi ditentukan dari titik-titik tersebut (Wahyono et al., 2020). Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya 14 paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Algoritma KNN termasuk metode yang menggunakan algoritma *supervised*. Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data (Liantoni, 2016). Cover dan Hart memperkenalkan KNearest Neighbor pada tahun 1968. K-Nearest Neighbor adalah metode klasifikasi pemalas karena algoritma ini menyimpan semua nilai data pelatihan dan menunda proses pembentukan model klasifikasi hingga data uji diberikan untuk prediksi (Mulak & Talhar, 2013). Penerapan algoritma K-Nearest Neighbor akan penulis lakukan dengan pengambilan data pada sistem akademik di Universitas PGRI Mahadewa Indonesia, kemudian akan menjadikan sebagian data menjadi data latih dan sebagian lagi menjadi data uji yang akan diproses dengan algoritma K-Nearest Neighbors untuk menentukan kelulusan tepat atau kelulusan terlambat mahasiswa. Penulis menggunakan Bahasa pemrograman Python dan menggunakan aplikasi pendukung yaitu Visual Studio Code dan Streamlit, Bahasa Pemrograman Python diciptakan oleh Guido van Rossum di Belanda pada tahun 1990 dan namanya diambil dari acara televisi kesukaan Guido Monty Python's Flying Circus. Van Rossum mengembangkan Python sebagai hobi, kemudian Python menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintak intuitif dan memiliki putaka yang luas. Python adalah bahasa pemrograman interpretatif yang dianggap mudah dipelajari serta berfokus pada keterbacaan kode. Python secara umum berbentuk pemrograman berorientasi objek, pemrograman imperatif dan pemrograman fungsional. Fitur dan kelebihan python, yaitu: 1. Memiliki koleksi kepastakaan yang banyak, tersedia modul-modul yang 'siap pakai' untuk berbagai keperluan; 2. Memiliki struktur bahasa yang jelas, sederhana, dan mudah dipelajari. 3. Berorientasi objek; 4. Memiliki sistem pengelolaan memori otomatis; 5. Bersifat modular (Romzi & Kurniawan, 2020). Visual Studio Code adalah aplikasi kode editor buatan Microsoft yang dapat dijalankan di semua perangkat desktop secara gratis. Kelengkapan fitur dan ekstensi membuat kode editor ini menjadi pilihan utama para pengembang. Visual Studio Code bahkan mendukung hampir semua sistem operasi seperti Windows, Mac OS, Linux, dan lain sebagainya. Visual Studio

Code adalah editor kode sumber yang ringan namun kuat yang berjalan di desktop dan tersedia untuk Windows, macOS, dan Linux. Muncul dengan dukungan *built-in* untuk JavaScript, TypeScript dan Node.js dan memiliki ekosistem ekstensi yang kaya untuk bahasa lain (seperti C ++, C #, Java, Python, PHP, Go) dan *runtime* (seperti .NET dan Unity). *Paper* ini dimaksud untuk memudahkan dalam pembelajaran dasar pemrograman Python menggunakan visual studio code yang dimulai dari instalasi python, instalasi visual studio code pada komputer, serta pemrograman python pada visual studio code berdasarkan algoritma pemrograman (Romzi & Kurniawan, 2020). Streamlit adalah kerangka kerja pengkodean Python *open-source* untuk membangun aplikasi web atau "aplikasi web" dan sekarang digunakan oleh para peneliti untuk berbagi kumpulan data besar dari studi yang diterbitkan dan sumber daya lainnya (Nápoles-Duarte et al., 2022).

Dengan adanya proses sesuai yang disebutkan diatas, diharapkan penerapan K-Nearest Neighbors dengan melakukan Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neighbors (KNN) Studi Kasus Universitas PGRI Mahadewa Indonesia dapat diprediksi dan membantu pengguna.

## **METODE**

### 1. Objek Penelitian

Objek dari penelitian ini adalah untuk melakukan Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neighbors (KNN) di Universitas PGRI Mahadewa Indonesia.

### 2. Metode Pengumpulan Data

Metode yang diperlukan untuk mengumpulkan data dan mengolah data yang di dapat dari objek sehingga diharapkan penelitian ini berjalan dengan baik dan sistematis.

Adapun pengumpulan data dilakukan dengan cara :

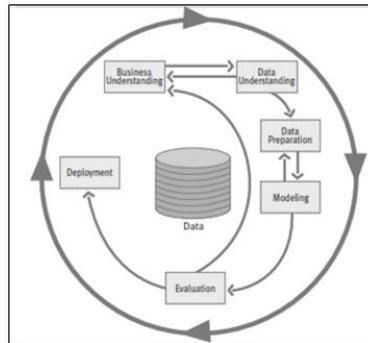
#### a. Observasi

Metode pengumpulan data dengan melakukan pengamatan. Observasi ini sangat penting, dimana penelitian akan melakukan pengamatan pada sumber data yang akan di dapat. Maka penulis melakukan observasi langsung dengan mengambil data kelulusan mahasiswa di Universitas PGRI Mahadewa Indonesia.

#### b. Studi Literatur

Penulis melakukan studi pustaka pada buku-buku, jurnal, pengumpulan data, dan referensi yang mendukung laporan penelitian ini.

### 3. Metodologi CROSS-INDUSTRY STANDARD FOR DATA MINING (CRISP-DM)



Gambar 1 CRISP-DM

Sumber : (Hidayat Zulfallah, 2022)

#### a. *Business Understanding*

Pada fase ini perlu dipahami pentingnya penggunaan data mahasiswa, sehingga dapat digunakan untuk menentukan kelulusan mahasiswa berdasarkan data yang diperoleh. Dengan melakukan observasi ditempat penelitian akan dilakukan yaitu Universitas PGRI Mahadewa Indonesia. Maka kita akan memperoleh data yang di perlukan untuk penelitian, data tersebut harus memberikan informasi untuk mencapai pemahaman bisnis terkait keakuratan kelulusan di Universitas PGRI Mahadewa Indonesia.

#### b. *Data Understanding*

Pada fase ini dilakukan pengumpulan data yang di dapat dari data kelulusan mahasiswa angkatan tahun 2022 Universitas PGRI Mahadewa Indonesia yang akan lebih di pahami lagi. Fase ini terdiri dari memahami data yang di peroleh untuk mencari informasi yang dapat diperoleh serta mengevaluasi kualitas dan kelengkapan data.

#### c. *Data Preparation*

Pada fase ini akan dilakukan pemilihan kriteria data yang akan digunakan dengan cara menentukan atribut dan labelnya, serta menghapus atribut data yang tidak digunakan.

#### d. *Modeling*

Pada fase ini akan dilakukan proses prediksi menggunakan algoritma K-Nearest Neighbors menggunakan aplikasi Visual Studio Code dan bahasa pemrograman Python.

#### e. *Evaluation*

Melakukan interpretasi terhadap hasil dari data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya. Berdasarkan tujuan yang di jelaskan pada fase *business understanding*, maka pada fase ini kineja dari algoritma yang digunakan diukur dengan melihat nilai accuracy dan precision.

#### f. *Deployment*

Pada fase ini hasil yang diperoleh dari penelitian ini berupa laporan hasil analisis data prediksi kelulusan mahasiswa di Universitas PGRI Mahadewa Indonesia yang diperoleh apakah berjalan dengan baik.

## HASIL DAN PEMBAHASAN

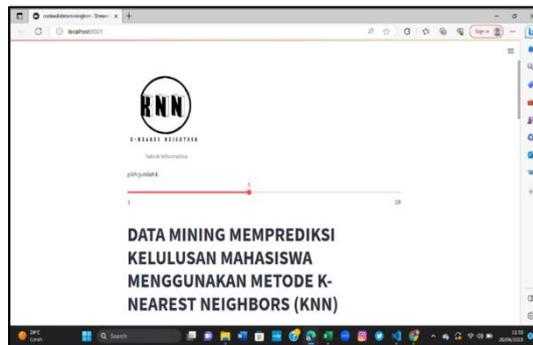
### 1. Implementasi Model

Dalam pengimplementasian model tahapan awal yaitu *business understanding* yang dimiliki seperti yang dijelaskan pada bab 3, kemudian lanjut ke tahap data *understanding* yaitu tahap mempelajari data, setelah itu dilanjutkan lagi ke tahap data *preparation* yaitu tahap memperbaiki data dan menentukan kriteria data yang akan digunakan dengan cara menentukan atribut dan labelnya, serta menghapus atribut data yang tidak digunakan, kemudian ke tahap *modeling* pada tahap ini akan dilakukan proses prediksi dengan algoritma K-Nearest Neighbors menggunakan aplikasi Visual Studio Code dan bahasa pemrograman Python, kemudian dilanjutkan ke tahap *evaluation* yaitu tahap memvalidasi hasil dari pemodelan di tahap sebelumnya, dan yang terakhir adalah tahap *deployment* yaitu pada tahap ini hasil yang diperoleh dari penelitian ini berupa laporan hasil analisis data prediksi kelulusan mahasiswa di Universitas PGRI Mahadewa Indonesia yang diperoleh apakah berjalan dengan baik. Sebelum masuk ke dalam proses pemodelan, data sudah dilakukan perbaikan dan pemilihan kriteria secara manual di proses persiapan. Kriteria yang digunakan adalah jenis kelamin, IPK, tahun masuk, tahun lulus, dan status.

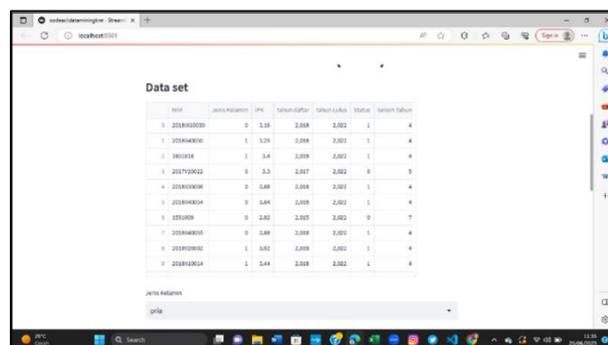
### 2. Pembahasan

Berdasarkan model yang telah dibuat, berikut implementasi aplikasi data mining prediksi kelulusan menggunakan metode K-Nearest Neighbor sebagai berikut.

#### a. Hasil Implementasi Aplikasi

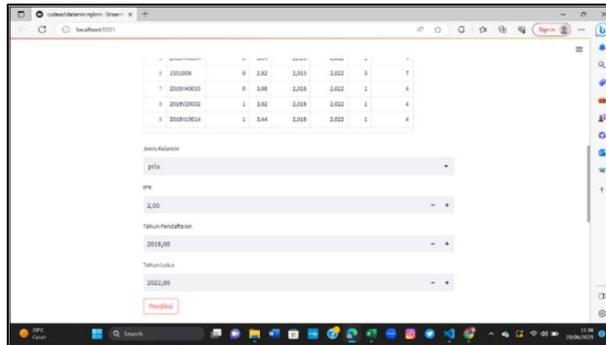


Gambar 2 Tampilan Antarmuka Pemilihan

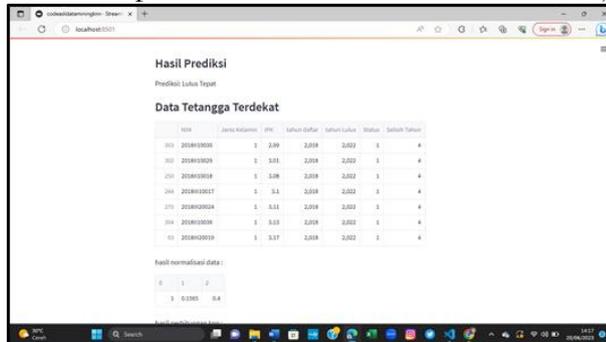


id	jenis kelamin	IPK	tahun masuk	tahun lulus	status	kelulusan	
0	2018010001	0	2,89	2,008	2,002	1	4
1	2018040001	1	3,19	3,008	3,002	1	4
2	19030001	1	3,4	3,008	2,002	1	4
3	2017030022	0	3,2	3,002	2,002	0	0
4	2018030004	0	3,88	3,008	2,002	1	4
5	2018080004	0	3,84	3,008	2,002	1	4
6	19050001	0	2,82	2,002	2,002	0	7
7	2018080005	0	3,88	3,008	2,002	1	4
8	2018020002	1	3,82	3,008	2,002	1	4
9	2018030004	1	3,44	3,008	2,002	1	4

Gambar 3 Tampilan Antarmuka Data Set

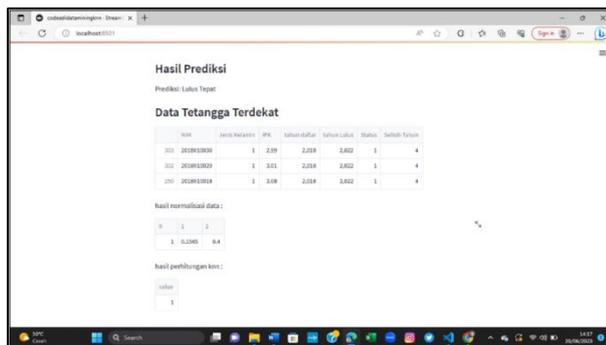


Gambar 4 Tampilan Antarmuka Memasukan Data Uji Baru



Gambar 5 Tampilan Antarmuka Hasil Prediksi

- b. Pembahasan prediksi dengan K yang berbeda  
 Melakukan prediksi menggunakan data baru dengan format jenis kelamin : pria, IPK : 3,00, tahun daftar : 2018, tahun lulus : 2022, dan k : 3. Dengan hasil prediksi dapat dilihat pada gambar 6.



Gambar 6. Hasil Prediksi dengan K =3

Melakukan prediksi menggunakan data baru dengan format jenis kelamin : pria, IPK : 3,00, tahun daftar : 2018, tahun lulus : 2022, dan k : 5. Dengan hasil prediksi dapat dilihat pada gambar 7

**Hasil Prediksi**  
Prediksi Lulus Tepat

**Data Tetangga Terdekat**

NIM	Jenis Kelamin	IPK	tahun daftar	tahun Lulus	Status	Selisih Tahun
100	2018H0000	1	2,09	2,019	2,022	3
100	2018H0009	1	2,01	2,019	2,022	3
100	2018H0018	1	2,09	2,019	2,022	3
104	2018H0017	1	3,4	2,019	2,022	3
170	2018H0014	1	3,11	2,019	2,022	3
104	2018H0006	1	3,13	2,019	2,022	3
10	2018H0010	1	3,17	2,019	2,022	3

hasil normalisasi data:

1	2
3	0,3365

Gambar 7. Hasil Prediksi dengan K =5

Melakukan prediksi menggunakan data baru dengan format jenis kelamin : pria, IPK : 3,00, tahun daftar : 2018, tahun lulus : 2022, dan k : 7. Dengan hasil prediksi dapat dilihat pada gambar 8 .

**Hasil Prediksi**  
Prediksi Lulus Tepat

**Data Tetangga Terdekat**

NIM	Jenis Kelamin	IPK	tahun daftar	tahun Lulus	Status	Selisih Tahun
101	2017H0010	1	2,99	2,019	2,022	3
101	2017H0029	1	3,61	2,019	2,022	3
101	2017H0062	1	3,05	2,017	2,022	5
101	2017H0018	1	3,08	2,019	2,022	3
104	2018H0017	1	3,1	2,019	2,022	3

hasil normalisasi data:

1	2
1	-4,753

Gambar 8. Hasil Prediksi dengan K =7

Dari ketiga prediksi yang ada dapat dilihat hasil prediksi adalah lulus tepat dengan jumlah tetangga yang berbeda-beda. Namun masih ada kemungkinan hasil akan berbeda jika data baru yang di masukan berbeda dari data yang digunakan saat ini. Sedangkan dengan menggunakan data tahun 2021 dengan k yang berbeda menggunakan jenis kelamin : pria, IPK : 3,00, tahun daftar : 2016, dan tahun lulus : 2021 hasilnya seperti pada gambar 9 dan 10 di bawah ini. Dari kedua prediksi yang ada dapat dilihat hasil prediksi adalah lulus terlambat dengan jumlah tetangga yang berbeda-beda.

**Hasil Prediksi**  
Prediksi Lulus Terlambat

**Data Tetangga Terdekat**

NIM	Jenis Kelamin	IPK	tahun daftar	tahun Lulus	Status	Selisih Tahun
55	2016H10072	1	3	2,016	2,021	0
47	2016H10043	1	3	2,016	2,021	0
57	2016H40054	1	2,97	2,016	2,021	0

hasil perhitungan k:

Gambar 9. Hasil Prediksi dengan K =3

**Hasil Prediksi**  
Prediksi: Lulus Terlambat

**Data Tetangga Terdekat**

	NIM	Jenis Kelamin	IPK	tahun daftar	tahun Lulus	Status	Selsih Tahun
55	2016R10072	1	3	2,016	2,021	0	5
47	2016R10043	1	3	2,016	2,021	0	5
57	2016R40054	1	2,97	2,016	2,021	0	5
54	2016R10067	1	2,96	2,016	2,021	0	5
63	2016R10056	1	3,04	2,016	2,021	0	5

Gambar 10. Hasil Prediksi dengan K =5

## SIMPULAN

Data Mining prediksi kelulusan menggunakan metode K-Nearest Neighbor dalam skripsi ini dalam pembuatannya melalui tahapan *business understanding*, kemudian lanjut ke tahap *data understanding* yaitu tahap mempelajari data, setelah itu dilanjutkan lagi ke tahap *data preparation* yaitu tahap memperbaiki data dan menentukan kriteria data yang akan digunakan dengan cara menentukan atribut dan labelnya, serta menghapus atribut data yang tidak digunakan, kemudian ke tahap *modeling* pada tahap ini akan dilakukan proses prediksi dengan algoritma K-Nearest Neighbors menggunakan aplikasi Visual Studio Code dan bahasa pemrograman Python, kemudian dilanjutkan ke tahap *evaluation* yaitu tahap mengevaluasi hasil dari pemodelan di tahap sebelumnya, dan yang terakhir adalah tahap *deployment*. Ada pun hasil analisis data prediksi kelulusan mahasiswa menggunakan metode K-Nearest Neighbor sebagai berikut : Dapat membantu mengetahui cara kerja algoritma K-Nearest Neighbor yaitu dengan melakukan prediksi target dari data baru yang diujikan pengguna. Pada data terdekat sesuai k yang dimasukkan, target terbanyak akan menjadi hasil dari prediksi, Dapat membantu pengguna dalam menentukan kelulusan mahasiswa, apakah mahasiswa lulus tepat atau lulus terlambat dengan melihat hasil dan data terdekat dari data yang diujikan, Mengetahui hasil prediksi dengan tetangga terdekat (k) yang berbeda.

Saran untuk peneliti selanjutnya yaitu Berdasarkan perancangan dan implementasi yang dilakukan. Penulis berharap sistem ini dapat dikembangkan oleh pengembang selanjutnya, dengan ruang lingkup yang dapat ditambahkan yaitu fitur SKS, Prodi, dan IPS serta menggunakan dataset yang baru. Sistem dapat juga dikembangkan kedalam platform lain seperti android dan ios.

## DAFTAR PUSTAKA

- Hidayat Zulfallah, F. (2022). SKRIPSI IMPLEMENTASI ALGORITMA KNN DALAM MENGUKUR KETEPATAN KELULUSAN MAHASISWA UIN SYARIF HIDAYATULLAH JAKARTA.
- Kadek, ( I, Arta, J., Indrawan, G., Gede, ), & Dantes, R. (2019). DATA MINING REKOMENDASI CALON MAHASISWA BERPRESTASI DI STMIK DENPASAR MENGGUNAKAN METODE TECHNIQUE FOR OTHERS REFERENCE BY SIMILARITY TO IDEAL SOLUTION. *Jurnal Ilmu Komputer Indonesia (JIKI)*, 4(1).
- KEMENDIKBUD. (2020). BERITA NEGARA REPUBLIK INDONESIA. [www.peraturan.go.id](http://www.peraturan.go.id)
- Liantoni, F. (2016). Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *Jurnal ULTIMATICS*, 7(2), 98–104. <https://doi.org/10.31937/ti.v7i2.356>
- Meilina, P. (2015). PENERAPAN DATA MINING DENGAN METODE KLASIFIKASI MENGGUNAKAN DECISION TREE DAN REGRESI. In Januari (Vol. 7, Issue 1).

- Mulak, P., & Talhar, N. (2013). Analysis of Distance Measures Using K-Nearest Neighbor Algorithm on KDD Dataset. In International Journal of Science and Research (Vol. 4). [www.ijsr.net](http://www.ijsr.net)
- Nápoles-Duarte, J. M., Biswas, A., Parker, M. I., Palomares-Baez, J. P., Chávez-Rojo, M. A., & Rodríguez-Valdez, L. M. (2022). Stmol: A component for building interactive molecular visualizations within streamlit web-applications. *Frontiers in Molecular Biosciences*, 9. <https://doi.org/10.3389/fmolb.2022.990846>
- Romzi, M., & Kurniawan, B. (2020). Implementasi Pemrograman Python Menggunakan Visual Studio Code. In JIK: Vol. XI (Issue 2). [www.python.org](http://www.python.org)
- Wahyono, W., Trisna, I. N. P., Sariwening, S. L., Fajar, M., & Wijayanto, D. (2020). Comparison of distance measurement on k-nearest neighbour in textual data classification. *Jurnal Teknologi Dan Sistem Komputer*, 8(1), 54–58. <https://doi.org/10.14710/jtsiskom.8.1.2020.54-58>