

Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama

(Studi Kasus : SMPN 3 Karawang Barat)

Desti Puspita Sari ^a, Shofa Shofia Hilabi ^b, Agustia Hananto ^{c,*}

^{a,b,c}Sistem Informasi, Universitas Buana Perjuangan

*si19.destisari@mhs.ubpkarawang.ac.id

shofa.hilabi@ubpkarawang.ac.id

agustia.hananto@ubpkarawang.ac.id

Abstract - Technology has become an important need for society that supports all activities to streamline student graduation with an average credit card, reduces problems related to predicting student graduation, the application of data mining to predict school graduation. The need for data mining stems from the amount of data that can be retrieved useful information and insights. Based on the results of calculations carried out using the orange modeling, the average grades of Class IX students vary from 80 to 90, and all students are declared PASS because their grades meet the graduation requirements set. is in SMPN 3 West Karawang. The K-Nearest Neighbor algorithm is useful for predicting a large number of graduates due to the consistent data processing of these predictions. For further research, use all student data to predict class growth towards student graduation..

Keywords - K-Nearest Neighbor; Prediction; National Examination.

Abstrak - Teknologi sudah menjadi kebutuhan penting bagi masyarakat yang mendukung segala aktivitas untuk mengefisienkan kelulusan mahasiswa dengan kartu kredit rata-rata, mengurangi permasalahan terkait prediksi kelulusan siswa, penerapan data mining untuk memprediksi kelulusan sekolah. Kebutuhan data mining bermula dari banyaknya data yang dapat diambil informasi dan wawasan yang berguna. Berdasarkan hasil perhitungan yang dilakukan dengan menggunakan pemodelan orange, nilai rata-rata siswa Kelas IX bervariasi dari 80 hingga 90, dan semua siswa dinyatakan LULUS karena nilai mereka memenuhi persyaratan kelulusan yang ditetapkan. berada di SMPN 3 Karawang Barat. Algoritma K-Nearest Neighbor berguna untuk memprediksi lulusan dalam jumlah besar karena pemrosesan data yang konsisten dari prediksi tersebut. Untuk penelitian selanjutnya, gunakan semua data siswa untuk memprediksi pertumbuhan kelas terhadap kelulusan siswa.

Kata Kunci— K-Nearest Neighbor; Prediksi; Ujian Nasional.

I. PENDAHULUAN

Dengan teknologi mutakhir saat ini berkembang Teknologi telah menjadi faktor penting dalam kehidupan kita sehari-hari., dan dengan cepat semakin terkait dengan kehidupan kita sehari-hari. Teknologi telah memberikan solusi yang dapat digunakan oleh semua orang yang mendukung semua kegiatan untuk membuat lebih efisien. Bahkan di zaman teknologi canggih ini, contoh tersisa dapat diprediksi dengan memanfaatkan teknologi yang tersedia, dengan akurasi 100% [1].

SMPN 3 Karawang Barat untuk saat ini telah menghapuskan UN (Ujian Nasional) oleh karena itu untuk menentukan kelulusan siswa menggunakan perhitungan nilai rata-rata pada raport, agar mengurangi permasalahan dalam memprediksi kelulusan siswa maka penulis memberikan saran melalui Tugas Akhir ini dengan mengusulkan Penerapan Data Mining Memprediksi Kelulusan Siswa Menggunakan Algoritma *K-Nearest Neighbour*[2]. Dikarenakan data yang dikelola saat ini oleh pihak sekolah tidak optimal dan sistem kelulusan pemerintah saat ini menciptakan ketidakpastian kelulusan siswa, membuat siswa dan guru resah. Dengan sistem diperbarui yang diperkenalkan oleh pemerintah ini, sulit bagi sekolah untuk mencoba memperbaiki kelulusan siswanya. Tanpa sistem prediksi, minimnya kelulusan siswa sangat tinggi karena pihak sekolah tidak tanggap terhadap masalah tersebut.

Untuk mengatasi ini, anda memerlukan sistem menerapkan ilmu data mining untuk memprediksi keberhasilan UN (Ujian Nasional). Kebutuhan akan data mining muncul dari data besar yang benar-benar dapat diterapkan dimanfaatkan untuk mengekstrak informasi dan wawasan. Informasi dan data dapat digunakan untuk mengumpulkan dan menganalisis sejumlah besar data serta untuk mengumpulkan informasi tentang fakta yang sebelumnya tidak diketahui[3].

II. METODE

A. Data Mining

Penambangan data adalah disiplin ilmu yang melibatkan penemuan pola menarik dari sejumlah data penting[4]. Data dapat dimasukkan ke dalam database, spreadsheet, atau perangkat penyimpanan informasi lainnya.. Penambangan data terkait dengan disiplin ilmu lain seperti sistem basis data, pergudangan data, statistik, pembelajaran mesin, pencarian informasi, komputasi tingkat tinggi, dan pengenalan pola. Penambangan data juga melibatkan banyak ilmu lain seperti analisis data spasial, jaringan saraf, dan pengenalan pola.

B. Tahapan – Tahapan Data Mining

Penambangan data adalah proses tersendiri memiliki 6 fase. Langkah-langkah ini bersifat interaktif dan pengguna terlibat secara langsung atau melalui database sebagai berikut[5]:

1. *Cleaning Data* (Pembersihan Data) Pembersihan Data menghapus data yang benar-benar salah atau tidak diperlukan serta kebisingan. Umum, baik basis data perusahaan dan hasil tes memiliki entri yang tidak lengkap, lebih menyukai: B. data hilang, data salah atau hanya salah ketik. Selain itu, terdapat atribut data yang tidak sesuai dengan hipotesis data mining itu sendiri. Juga lebih baik untuk membuang informasi yang tidak relevan. Pembersihan data juga memengaruhi kinerja teknik penambangan data dengan mengurangi kuantitas dan kompleksitas data yang akan dianalisis[6].
2. *Data Integration* (data integritas) Integrasi data adalah proses mentransfer informasi database. ke basis data lain. Informasi yang dibutuhkan untuk penambangan data atau lebih database, tetapi begitu banyak kumpulan data atau file [5]. Integrasi data dilakukan untuk atribut yang memungkinkan identifikasi entitas unik, seperti atribut B. Nama, jenis produk, dan lain-lain. Data terintegrasi harus selalu dilakukan dengan hati-hati karena ada risiko yang terkait. dapat menyebabkan hasil yang tidak jelas di masa mendatang.
3. Pemilihan Data sering di database tidak semuanya digunakan, sehingga hanya data yang cocok untuk analisis saja yang dicari di database. Misalnya, dalam kasus memeriksa keterampilan membeli orang, Anda tidak harus menggunakan nama pelanggan dalam analisis keranjang belanja, ID pelanggan sudah cukup[7].
4. *Data Transformation* (transformasi data) Data diubah atau dimasukkan bentuk yg sinkron buat diproses pada data mining [8]. Beberapa metode penambangan data memerlukan penggunaan format data sebenarnya. Faktanya, banyak teknik standar seperti analitik asosiasi dan kelompok hanya dapat menerima input kategori untuk data. Untuk alasan ini, data harus dibagi menjadi beberapa interval waktu dalam bentuk angka numerik yang berurutan.
5. Proses penambangan adalah Proses primer yg memakai metode buat menemukan keterangan berharga & tersembunyi pada data [9].
6. *Model Evaluation* (Model Evaluasi), menemukan model yang menarik atas informasi yang ditemukan. Pada fase ini, output berdasarkan teknik data mining dinilai pada bentuk pola tipikal & contoh prediksi buat menilai apakah hipotesis yg terdapat telah terpenuhi[10]. Jika hasil yang diperoleh ternyata tidak akurat dengan hipotesis, unsur-unsur tertentu pilihan, misalnya. menggunakannya sebagai Umpan balik untuk mempercepat prosedur pengolahan data, untuk mencoba metode penambangan data lain yang lebih sesuai atau untuk mendapatkan hasil tersebut. sebagai hasil yang tidak terduga dapat bermanfaat.

C. Pengertian Prediksi

Mencoba meminimalkan perbedaan antara perkiraan dan hasilnya disebut akurasi. Seorang peramal perlu mengumpulkan informasi dari tren masa lalu dan sekarang untuk membuat prediksi yang akurat. Orang perlu menyadari bahwa ramalan tidak boleh memberikan jawaban pasti tentang apa yang terjadi. Namun, peramal dapat menggunakan pengetahuannya untuk mendekati hasil sedekat mungkin.[11]

D. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor, atau *K-NN*, adalah metode yang menggunakan algoritma terawasi. Prinsip kerja *KNN* adalah mengklasifikasikan data baru berdasarkan atribut dan contoh yang digunakan dalam pelatihan[12]. *KNN* menggunakan k tetangga dalam data pelatihan sebagai titik awal untuk kueri. Tujuan dari *KNN* adalah mengklasifikasikan data baru dengan mencari jarak terpendek antara data baru dengan k tetangganya. Berdasarkan kajian *KNN*, hasil tes dibagi menjadi kategori berdasarkan jawaban terbanyak. Ini ditentukan dengan menggunakan *Euclidean Distance*. Pengukuran jarak yang paling umum digunakan adalah *Euclidean*. Setiap pengukuran memiliki definisi tertentu. Definisi pengukuran *Euclidean* adalah sebagai berikut[13]:

$$D(x,y) = \sqrt{\sum_{k=1}^n (X_{training} - Y_{testing})^2} \quad (1)$$

- X_{training} : data training ke-i
- Y_{testing} : data testing
- i : record (baris) ke-i dari tabel
- n : jumlah data training

III. HASIL DAN PEMBAHASAN

A. Analisa Masalah

Analisis masalah saat ini merupakan gambaran dari sistem yang ada digunakan saat ini, dimaksudkan untuk mengidentifikasi dan mengevaluasi sistem memprediksi kelulusan siswa sekolah menengah pertama[14]. Berikut adalah data sampel yang digunakan oleh penulis adalah siswa kelas IX :

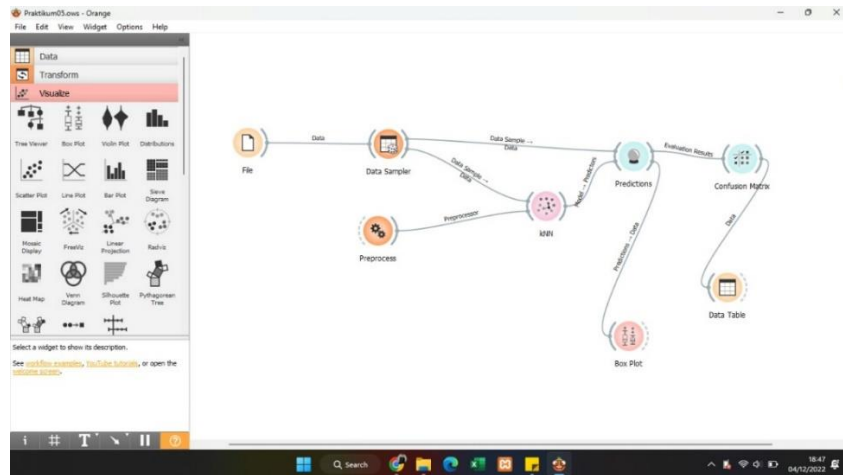
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	NISN	NAMA SISWA	paik	pkn	b.indo	mtk	ipa	ips	b.ling	sbk	PJOK	Prakerja	b.sund	Rata-Rata				
1	0068466024	AA ISYAH SHABIYAHKI	91	94	93	90	93	94	94	94	90	94	94	92.8182				
2	0071954897	ADAM AIRLANGGA PRA	92	94	93	93	93	93	90	94	92	94	93	92.8182				
3	0064602374	ADITYA RAHMAY SANTI	92	93	92	90	92	92	89	94	90	93	93	91.8182				
4	0073370393	ANDIKA APRILLAN	88	91	91	87	90	91	87	93	88	89	91	89.6364				
5	0056090333	ANDRI AGUSTIAN	88	91	90	86	89	91	87	92	90	89	91	89.4545				
6	0073024742	ANDRIANSYAH	88	92	90	87	89	91	87	93	88	89	91	89.5455				
7	0073110446	ARTATI SHARAHADHIPA	91	92	92	90	90	91	87	93	90	93	92	91				
8	0078159544	AURIA ALCHA KOSASIH	91	92	92	89	89	92	89	93	88	93	92	90.9091				
9	0077227104	AZZAHRA NUR FAHIRA	92	92	93	90	92	93	93	94	91	93	93	92.3636				
10	0077280793	BAYU CANDRA YUDISTI	91	92	92	89	91	92	88	93	91	92	92	91.1818				
11	0074974664	CANDRA SEPTIAN RIWA	87	88	89	87	88	88	86	91	88	91	89	88.2626				
12	0068974811	DESTA ALFIANA	87	88	90	85	89	90	86	92	86	91	89	88.4545				
13	0067479218	DESY ARUMSARI	92	90	91	88	89	90	88	93	89	91	92	90.2727				
14	0079179782	DWI INTAN VINCA ROSS	90	93	91	88	89	94	89	94	88	93	92	91				
15	0075747888	FADIL FIRMANSYAH ALI	88	91	89	86	88	91	86	92	89	91	92	89.3636				
16	0075962929	FATHIN NUR ROHMANI	91	92	89	89	89	90	88	92	89	92	91	90.1818				
17	0074354119	FELISHA HADIVA	92	94	91	90	89	92	89	94	89	92	92	91.2727				
18	0063370972	GAGAH PANJLI NUGRAH	88	92	92	87	90	92	89	93	90	92	92	90.6364				
19	0068145303	GITA	93	90	92	90	90	91	88	93	89	94	91	91				
20	0051265736	HALISRA ALDAPFA	91	93	92	91	92	92	90	94	92	94	92	92.0909				
21	0068609397	INTAN DZAKIRA	89	91	91	87	88	91	88	94	88	91	91	89.9091				
22	0075271778	KAYLA CHOERUNNISA	92	93	92	90	91	92	90	93	89	92	91	91.3636				
23	0075271778	KAYLA CHOERUNNISA	89	91	91	89	90	92	88	93	89	93	92	90.6364				
24	0079375024	KIRYATI TRIS FEBRIANI	91	91	91	88	88	91	88	93	88	93	92	90.2727				

Gambar 1. Data Siswa Kelas IX

B. Penyelesaian Orange :

1. Langkah pertama pemodelan data mining

Dan langkah selanjutnya adalah memodelkan proses menggunakan data testing yang telah dipilih tadi ke dalam aplikasi *Orange*. *K-Nearest Neighbour* kemudian disambungkan *tools prediction* dan *confus matrix*.



Gambar 2. Testing Pada Orange

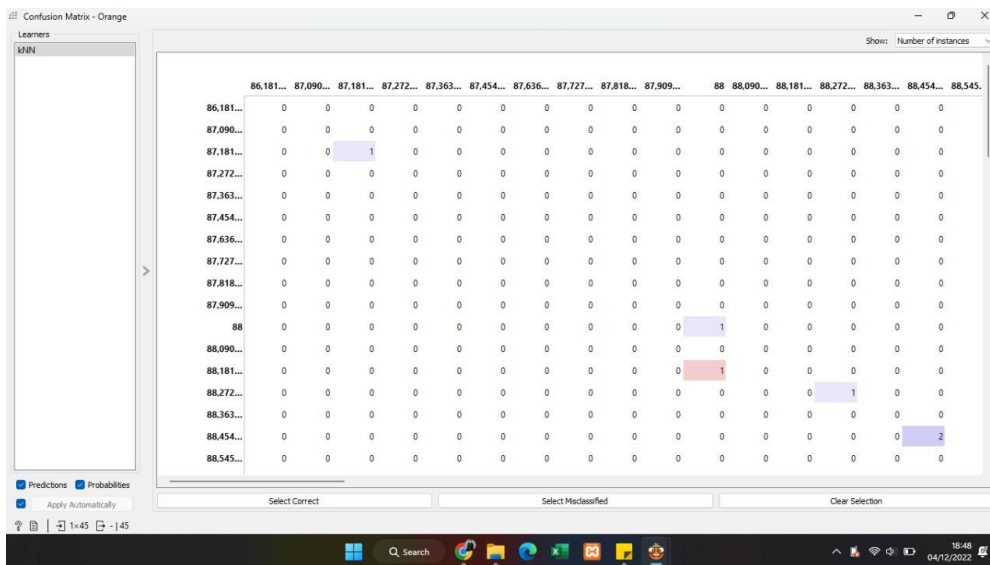
Berikut hasil pemodelan menggunakan orange :

1. hasil perhitungan Orange dengan metode *K-Nearest Neighbour*

	Rata-Rata	NAMA SISWA	pa1 Ijazah	pa2 Ijazah	pa3 Ijazah	mtk Ijazah	ipa Ijazah	ips Ijazah	ling Ijazah	sbk Ijazah	PJOK Ijazah
1	92,45454545	EMBUN ZATAS...	93	93	94	92	94	90	95	90	90
2	93,63636364	HAWIRA INSYI...	94	95	93	94	94	94	93	91	91
3	90,27272727	REKA RAHELI	88	90	89	86	88	88	89	87	89
4	91,45454545	AFFI NABIL DA...	87	90	89	88	88	89	87	91	87
5	93,63636364	REVANA FEBRI...	90	91	90	86	89	91	88	92	88
6	90,90909091	BIMA HARDIAL...	90	91	91	91	91	91	88	94	89
7	91,45454545	FARIH DWI AZA...	90	88	86	90	89	89	88	92	89
8	89,63636364	HIKMAH OKTA...	89	91	89	87	89	91	88	92	89
9	87,18181818	KUSNADI	88	87	88	85	86	88	85	90	86
10	88	JAVANA ABDI	88	87	86	87	89	86	87	92	87
11	91	NAHDA AQILA ...	91	92	91	90	90	91	91	93	88
12	90,90909091	SURYA AJINAJA	91	89	89	85	88	90	88	93	88
13	91,90909091	DOHA DODIA ...	90	91	90	90	88	90	91	91	88
14	90,90909091	ALLIFA NUR AZ...	91	91	91	90	89	91	90	91	88
15	92,45454545	NASYITHAH IR...	94	94	93	94	93	93	96	93	90
16	90,90909091	CANDRA BUDL...	89	88	87	88	89	88	88	92	88
17	90,72727273	RIAN PUTRA LA...	88	90	86	88	88	90	86	93	86
18	89,18181818	SRI AGUSTIN	88	91	89	90	88	91	88	93	88
19	90,27272727	SEPTIA WULAN...	92	90	90	89	89	90	89	92	88
20	88,27272727	WAHYU RAHM...	87	89	87	90	89	87	87	92	86
21	91,18181818	BAYU CANDRA ...	91	92	92	89	91	92	88	93	91
22	89,18181818	NOURUL ADIA...	88	90	88	87	88	89	89	92	89
23	90	FALZAN DRIN...	90	92	90	86	90	91	89	93	90
24	90,18181818	TIARA AJELIA R...	88	92	91	91	90	91	88	93	88
25	89,18181818	NAIFA HANNID...	88	91	91	86	89	91	88	93	88
26	91	RESSA FAYNAR	92	92	90	91	90	90	88	93	89
27	90,72727273	MOCH CANDR...	91	91	91	91	90	91	88	94	88
28	89,72727273	APRIYANTO	89	91	89	88	89	92	87	95	88
29	90,72727273	ROHATI NUR H...	91	91	91	90	91	91	91	92	87

Gambar 3. Hasil Perhitungan pada Orange

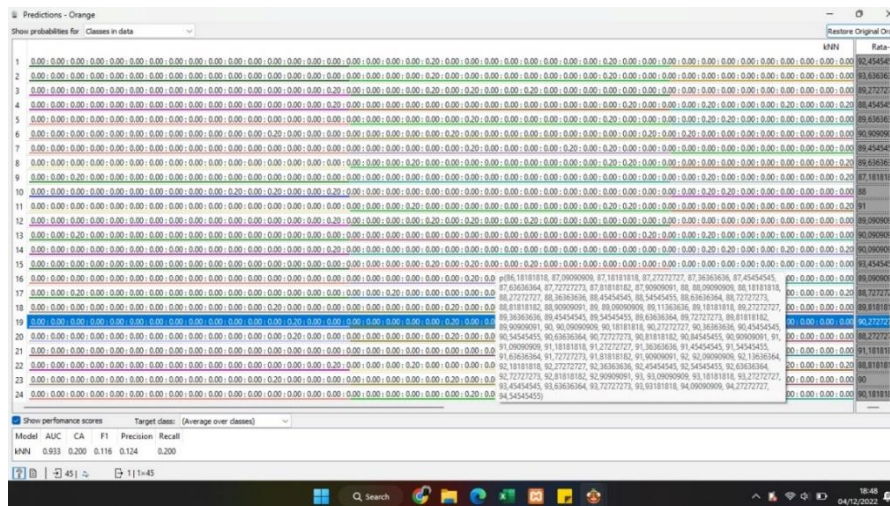
2. Gambar perhitungan *confus matrix*



Gambar 4. Perhitungan Confus Matrix

3. Hasil dari *prediction*

Dari hasil perhitungan menggunakan pemodelan orange rata-rata nilai siswa kelas IX adalah range 80 – 90, dan semua siswa dinyatakan LULUS karena nilai memenuhi standar kelulusan yang ada di SMPN 3 Karawang Barat.



Gambar 5. Hasil Perhitungan Prediction

IV. KESIMPULAN

Dari hasil penelitian menggunakan algoritma *K-Nearest Neighbour* yang penulis lakukan adalah :

1. Menghitung algoritma *K-Nearest Neighbour* ini menentukan hasil prediksi nilai kelulusan siswa rata-rata mencapai range 88 – 90, yang mana siswa tersebut dinyatakan LULUS karena nilai memenuhi standar.
2. Algoritma *K-Nearest Neighbour* cocok hasil untuk prediksi banyaknya jumlah siswa yang akan lulus, dikarenakan pengolahan data yang konsisten dari prediksi tersebut.

UCAPAN TERIMA KASIH

Saya mengirimkan ucapan terima kasih yang hangat kepada mereka seluruh anggota SMPN 3 Karawang Barat memberikan baik telah diberitahu bahwa saya penelitian di sekolah tertentu. Agar penelitian ini dapat berjalan lancar dan baik.

DAFTAR PUSTAKA

- [1] Gelinas, Ulric, Oram, Alan, Wiggins, and William, "Accounting Information System," pp. 17–30, 1990.
- [2] J. Manajemen, "Dirgamaya," vol. 02, no. 02, pp. 25–34, 2022.
- [3] A. Lia Hananto, B. Priyatna, and A. Haris, "Application of Prototype Method on Student Monitoring System Based on WEB," *Buana Inf. Technol. Comput. Sci. (BIT CS)*, vol. 1, no. 1, pp. 1–4, 2020, doi: 10.36805/bit-cs.v1i1.683.
- [4] Tumini and L. Damayanti, "Aplikasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naïve Bayes Studi Kasus Smp Negeri 11 Kotabumi Utara," *Inform. SIMANTIK*, vol. 3, no. 2, pp. 23–30, 2018.
- [5] I. Surya, "Kebutuhan Perangkat Lunak Untuk Aplikasi Data Mining," *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 233–240, 2017, doi: 10.25077/teknosi.v3i2.2017.233-240.
- [6] A. M. Habibi and R. R. Santika, "Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Jurusan Menggunakan Metode Euclidean Distance Berbasis Web Pada SMP Setia Gama," *Skanika*, vol. 3, no. 4, pp. 7–14, 2020.
- [7] R. Pratama, B. Huda, E. Novalia, and H. Kabir, "Perbandingan Algoritma C4 . 5 dan Naïve Bayes dalam Menentukan Persediaan Stok," pp. 115–122, 2022, doi: 10.47002/metik.v6i2.379.
- [8] F. Mumpuni, D. Maharani, A. L. Hananto, S. S. Hilabi, and F. N. Apriani, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor," 2022, doi: 10.47002/metik.v6i2.372.
- [9] A. M. Siregar, "Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning," *Petir*, vol. 13, no. 2, pp. 138–147, 2020, doi: 10.33322/petir.v13i2.998.
- [10] T. Tukino and B. Huda, "Penerapan Algoritma K-Means Untuk Mendukung Keputusan Dalam Pemilihan Tema Tugas Akhir Pada Prodi Sistem Informasi Universitas Buana Perjuangan Karawang.," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–10, 2019, doi: 10.36805/technoxplore.v4i1.542.
- [11] M. S. Mustafa and I. W. Simpen, "Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Diponegara Makassar)," *Creat. Inf. Technol. J.*, vol. 1, no. 4, p. 270, 2015, doi: 10.24076/citec.2014v1i4.27.
- [12] A. Sanjaya, T. Wahyana, F. Sains, U. Panca, S. Bekasi, and C. Autor, "WEBSITE " Penerapan Metode K- Nearest Neighbour Untuk Sistem Prediksi Kelulusan Siswa Keterangan :," vol. 3, no. 1, pp. 31–47, 2022.
- [13] A. Yandi Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm," *Techno.Com*, vol. 17, no. 4, p. 9, 2018.
- [14] S. S. Hilabi and . P., "ANALISIS KEPUASAN PENGGUNA TERHADAP LAYANAN APLIKASI MEDIA SOSIAL WhatsApp MOBILE ONLINE," *Buana Ilmu*, vol. 3, no. 1, pp. 119–136, 2018, doi: 10.36805/bi.v3i1.461.
- [15] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

Desti Puspita Sari. mahasiswa semester VII (tujuh) Sistem Informasi Fakultas Teknik Ilmu Komputer Universitas Buana Perjuangan Karawang, lahir di Karawang pada tanggal 19 Desember 1998. Anak Tunggal dari pasangan bapak Dede Suryadi dan Ibu Yati Nurhayati. Pendidikan yang telah ditempuh Sekolah Menengah Pertama di SMPN 2 Karawang Barat pada tahun 2010 dan selesai tahun 2013. Pendidikan Sekolah Menengah Kerujian di SMK Rosma Karawang dengan jurusan Teknik Komputer dan Jaringan selesai pada tahun 2016. Pada tahun 2019 sampai saat ini penulis melanjutkan pendidikannya pada salah Perguruan Tinggi Swasta di Karawang yaitu Universitas Buana Perjuangan Karawang. Lalu, pada tahun 2020 – 2021 penulis pernah bergabung di Badan Eksekutif Mahasiswa (BEM) Fakultas Teknik dan Ilmu Komputer.

Shofa Shofia Hilabi, S.Kom., M.Kom, lulusan S1 tahun 1990, di Universitas Budi Luhur. Lalu, lulus S2 pada tahun 2014 di STMIK LIKMI. Saat ini penulis menjadi Dosen Prodi Sistem Informasi di Universitas Buana Perjuangan Karawang.

Agustia Hananto, S.Kom., M.Kom, lulusan S1 tahun 2016, di Sekolah Tinggi Teknologi Informatika Sony Sugema. Lalu, lulus S2 pada tahun 2021 di Universitas Budi Luhur. Saat ini penulis merupakan Dosen Prodi Sistem Informasi di Universitas Buana Perjuangan Karawang.