

PENERAPAN K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI HASIL PANEN KEDELAI BERDASARKAN DATA CUACA DAN TANAH

Raihan Ali Najmi¹

¹Ilmu Komputer, Universitas Pamulang Tangerang Selatan, Indonesia

Email :raihaanajmi2007@gmail.com

ABSTRAK: Kedelai merupakan bahan pokok yang dibutuhkan masyarakat pada setiap tahunnya. Fakta pada lapangan, para petani di Indonesia mengalami penurunan hasil produksi kedelai. Hal tersebut dikarenakan ketidakmampuan para petani dalam meliputi kualitas dan kuantitas kedelai pada saat masa panen. Solusi dari masalah tersebut adalah menggunakan klasifikasi K-Nearest Neighbor untuk mendapatkan masa panen kedelai berkualitas dan berkuantitas tinggi. Algoritma dari penelitian ini menggunakan K-Nearest neighbor menggunakan pre-processing manualisasi min-max. Manualisasi diperlukan untuk menyamakan jarak antara nilai satu fitur dengan nilai fitur lainnya, pada penelitian ini akan menggunakan nilai yang hanya berjarak antara 1 sampai 0. Setelah semua nilai pada data latih maupun data uji melewati fase manualisasi, maka baru bisa masuk ke proses Klasifikasi menggunakan K-NN. Pada proses K-NN nilai dari fitur data latih akan dihitung jaraknya terhadap data uji menggunakan euclidian distance, diurutkan, dan kemudian diambil sejumlah nilai K teratas. Nilai K yang akan digunakan dalam sistem ialah bernilai 5. Proses terakhir pada K-NN ialah mengambil voting untuk mendapatkan kelas baru data uji. Berdasarkan hasil pengujian yang telah dilakukan, algoritma dari sistem mendapatkan presentase keberhasilan sebesar 80% dengan nilai K paling optimal yaitu bernilai 5.

Kata kunci : Kedelai, Klasifikasi, K-Nearest

ABSTRACT:

Neighbor (KNN) Abstract Soybeans are an essential component that is required by the community on an annual basis. The fact is that farmers in Indonesia have seen a reduction in soybean production due to their inability to cover both the quality and quantity of soybeans during the harvest season. It has caused a decline in the overall quality of the soybean crop. Utilizing the K-Nearest Neighbor classification to achieve both high-quality and high-quantity soybean harvests is the solution to the difficulty that we are now facing. This research uses the K-nearest Neighbor algorithm, which incorporates human min-max preprocessing. For this study, employing values that are only between 1 and 0, manual adjustment of the distance between one feature value and another is required. Once all of the values in both the training data and the test data have been validated by hand, they are ready to move on to the next stage of the classification process, which is performed using K-NN. The feature value of the training data will be determined for its distance to the test data utilizing Euclidean length, sorted. Then the top K values will be considered in the K-NN process. The system will continuously operate with a value of 5 for the K variable. The voting step is the last in the K-NN process, which is used to obtain the new class of test data. According to the findings of the experiments that were carried out, the algorithm of the system was able to reach a success rate of 80%, with the value of K proving to be the most ideal.

Keywords : Soybeans, Classification, K-Nearest Neighbor

PENDAHULUAN

Kedelai di Indonesia sangat dibutuhkan untuk setiap tahunnya. Seiring dengan berjalannya waktu kebutuhan kedelai selalu meningkat setiap tahunnya. Hal ini dikarenakan kebutuhan kedelai secara nasional terus mengalami permintaan dari sisi konsumen. Fakta di lapangan ialah dari pihak petani kedelai di Indonesia sendiri justru mengalami penurunan Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 2102 Fakultas Ilmu Komputer, Universitas Brawijaya hasil produksi. Dalam penelitian (Sri Wahyuningsih, 2021) dipaparkan bahwa salah satu masalah kedelai di Indonesia ialah ketidakmampuan petani lokal untuk meningkatkan produksi kedelai. Ketidakmampuan ini meliputi kemampuan produksi dengan kuantitas yang dibutuhkan pasar dan juga ketidakmampuan petani dalam memenuhi kualitas yang tinggi. Masalah ini akan berefek juga pada harga kedelai di pasaran. tinggi. Harga tinggi justru akan makin memperburuk kondisi petani kedelai di Indonesia. Dalam penilitan (Nur Mahdi & Suharno, 2019) dipaparkan bahwa selama ini cara Indonesia dalam menghadapi kekurangan kedelai ialah menggunakan cara impor. Impor memang dapat menyelesaikan permasalahan kekurangan kuantitas kedelai, akan tetapi menambah permasalahan bagi petani lokal. Petani lokal yang tidak mampu untuk memenuhi keinginan pasar, sekarang harus dihadapkan dengan kenyataan bahwa mereka harus bisa memproduksi kedelai dengan kualitas yang baik. Beban yang bertambah pada petani mengakibatkan mereka lebih enggan untuk menanam kedelai. Dalam penelitian (Supadi, 2016) dipaparkan solusi yang baik bagi petani lokal ialah dengan meningkatkan kualitas kedelai. Harapan dengan adanya peningkatan kualitas ini dapat meningkatkan juga harga kedelai lokal di pasaran, sehingga bisa menekan impor yang berlebihan. Tahapan penanaman kedelai yang berkualitas tinggi dimulai pada bibit dan juga penentuan masa panen. Penentuan masa panen ini menjadi sangat penting dikarenakan setiap tumbuhan kedelai memiliki masa panen yang berbeda tergantung dari kualitas perawatan yang tercermin pada tumbuhan kedelai itu sendiri. Solusi yang ingin penulis berikan ialah sebuah penelitian, dimana penulis akan membuat sebuah sistem untuk mengetahui masa panen dari varietas kedelai. Hasil dari sistem ini akan memperlihatkan masa panen dari varietas unggul berdasarkan dari umur kedelai, rata hasil, tinggi tanaman, hama dan juga penyakit dari tumbuhan kedelai. Semua fitur untuk masukan sistem akan diinput secara manual oleh petani. Hal ini dikarenakan fitur yang dimasukkan ke dalam sistem membutuhkan keahlian petani untuk mendapatkan nilainya.

METODOLOGI PENELITIAN

- a. Penelitian ini dilakukan melalui beberapa tahapan:
Pengumpulan Data: Mengumpulkan data historis hasil panen kedelai, cuaca (suhu, curah hujan, kelembaban), dan karakteristik tanah.
- b. Pra-pemrosesan Data: Melakukan pembersihan data, normalisasi, dan pemisahan data pelatihan dan pengujian.
- c. Implementasi KNN: Menerapkan algoritma K-Nearest Neighbor untuk membangun model prediksi.
- d. Evaluasi Model: Mengukur akurasi dan performa model menggunakan metrik seperti RMSE dan MAE.

Metode yang akan dipakai dalam penelitian ini ialah K-Nearest Neighbor (K-NN). Dijelaskan pada penelitian (Farokhah, 2020) yang berfokus pada pengelompokan bunga menggunakan K-NN, mendapatkan hasil akurasi sebesar 90% pada k bernilai 5. Hasil akurasi pada penelitian tersebut menjadi landasan bahwa metode K-NN merupakan metode yang baik. Metode K-NN juga merupakan metode yang ringan sehingga metode tersebut sangat mendukung untuk pemrosesan yang cepat. K-Nearest Neighbor (KNN) adalah sebuah metode pengenalan pola paling sederhana, memiliki tingkat keberhasilan yang tinggi. KNN adalah sebuah classifier yang mudah, pada sampel data dapat diklasifikasikan berdasarkan kelas dari tetangga terdekatnya. KNN merupakan salah satu metode klasifikasi yang bertujuan untuk melakukan klasifikasi berdasarkan data yang sudah ditentukan. KNN merupakan metode supervised karena data barunya akan diklasifikasikan dengan jarak terdekatnya dengan data latih (Prahudaya & Harjoko, 2017). Solusi akhir yang penulis berikan untuk membantu para petani lokal Indonesia untuk

mendapatkan kedelai berkualitas tinggi ialah dengan membuat penelitian berjudul Klasifikasi Masa Panen Varietas Unggul Kedelai Menggunakan K-Nearest Neighbor (K-NN). Inputan dari sistem berupa nilai dari umur kedelai, rata hasil, tinggi tanaman, hama dan juga penyakit dari tumbuhan kedelai. Semua nilai tersebut akan dinormalisasi agar menghasilkan jarak antara fitur satu dengan fitur lainnya berdekatan, sehingga bisa membantu untuk proses selanjutnya. Semua fitur yang telah dinormalisasi akan masuk ke proses K-Nearest neighbor untuk mengetahui masa panen dari tumbuhan tersebut. Hasil dari sistem ini akan mengeluarkan tiga buah kelas yaitu umur genjah, umur sedang dan umur dalam.

Meliputi rancangan/model, tata cara teknik pengumpulan data/informasi, tempat dan waktu, serta proses pengolahan dan analisis data/ informasi. Di dalam Metoda tidak perlu menggunakan sub-sub judul.

PERANCANGAN DAN IMPLEMENTASI

Dalam bagian ini akan dipaparkan mengenai proses rancangan hingga implementatif dari sistem yang dikerjakan.

- a. Perancangan Sistem Perancangan sistem merupakan sebuah proses pembuatan algoritma yang nantinya diharapkan dapat menyelesaikan masalah yang ada. Algoritma yang akan dibuat berfokus pada dua buah sistem utama yaitu tahapan preprocessing dan tahapan klasifikasi menggunakan K-NN. Kedua tahapan tersebut akan saling memiliki keterkaitan, dimana preprocessing akan memberikan hasil yang lebih baik pada proses klasifikasi K-NN. Oleh karena itu, tahapan preprocessing akan diimplementasikan pada setiap data latih maupun data uji. Tujuan dari preprocessing untuk mempersiapkan data masa panen varietas unggul kedelai agar dapat diproses pada tahapan selanjutnya. Data latih yang telah didapatkan sebelumnya, akan diproses menggunakan normalisasi min-max. Dengan dilakukannya normalisasi tersebut, rentang nilai dari data latih akan menjadi seragam antara satu fitur dengan fitur lainnya. Rentang nilai yang akan digunakan dalam penelitian ini ialah 1-0. Berikut merupakan alur preprocessing dapat digambarkan pada gambar dibawah ini.



Gambar 1 . Diagram Alir Preprocessing Seacara perhitungan manual.

Dari Gambar 1. Diagram Alir Preprocessing Seacara perhitungan manual, preprocessing menggunakan normalisasi minmax diawali dengan mencari nilai minimal dan juga maksimal dari data latih. Setelah nilai MinMax dari setiap fitur pada data latih didapatkan, maka proses normalisasi dapat dihitung menggunakan rumus dibawah ini. $Normalisasi = \frac{(Data - Min)}{(Max - Min)} \times (AB)$ Dimana, Data = Data yang akan dinormalisasi Min = Nilai minimal dari fitur Max = Nilai Maksimal dari fitur A = Rentang nilai maksimum B = Rentang nilai minimum Proses selanjutnya setelah proses preprocessing dilakukan ialah masuk ke proses klasifikasi menggunakan K-NN. Pada tahap ini



Gambar 2 Diagram Alir Klasifikasi Metode KNN Secara perhitungan manual

Dari Gambar 2. Diagram Alir Klasifikasi Metode KNN Secara perhitungan manual, Klasifikasi K-NN dimulai dengan menghitung euclidian distance dari data uji ataupun data yang akan dicari kelasnya dengan data latih yang ada.

- b. Pengaruh Nilai K Terhadap Akurasi Prosedur yang akan dilakukan dalam pengujian ini ialah menjalankan kode program secara menyeluruh dengan memasukan data uji yang berbeda. Hasil yang dikeluarkan oleh sistem nantinya akan dibandingkan dengan hasil yang sebenarnya. Pada pengujian ini juga akan memakai nilai dari K yang berbeda, yaitu 3, 5, dan 7. Pengujian akan dilakukan sebanyak sepuluh kali perulangan. Harapan dari pengujian ini ialah didapatkan nya nilai K yang paling optimal untuk digunakan oleh sistem. Gambar 3. Tabel Hasil Pengujian Pengaruh Nilai K Terhadap Akurasi Berdasarkan pada tabel diatas, didapatkan presentase keberhasilan 70% untuk nilai K=3, 80% untuk nilai K=5, dan 50% untuk nilai K=7. Nilai K=7 mendapatkan presentase keberhasilan paling kecil dikarenakan dengan nilai K=7 sistem menerima data yang lebih banyak disaat akan melakukan voting, sehingga mengakibatkan data error masuk lebih banyak pula yang berdampak kelas yang salah yang memenangkan voting. Nilai K=3 yang menempati posisi kedua menunjukan bahwa data teratas yang didapatkan setelah pengurutan hasil euclidian distance dari terkecil ke terbesar ialah data yang memiliki kelas yang benar, tetapi masih memiliki sedikit kekurangan data untuk mencapai hasil yang optimal. Nilai K=5 memiliki jumlah data yang cukup, tidak terlalu sedikit dan tidak juga memiliki data yang berlebih untuk mendapatkan hasil yang baik. Kesimpulan dari pengujian ini didapatkan bahwa nilai K yang

optimal untuk digunakan oleh sistem ialah K bernilai 5 dengan presentase keberhasilan 80%.

Table 1 Hasil Data Penyakit

No	umur	Rata hasil	Tinggi tanaman	Ket hama	Ket penyakit	Kelas sebenar	Hasil
1.	80	1,5	55	3	3	Umur sedang	Umur sedang
2.	90	1,5	50	2	2	Umur sedang	Umur sedang
3.	90	1,5	50	1	1	Umur sedang	Umur sedang
4.	75	1,5	70	1	1	Umur genjah	Umur genjah
5.	90	1,5	70	1	1	Umur genjah	Umur genjah
6	60	1,7	75	1	1	Umur genjah	Umur genjah
7	90	3,2	52	1	2	Umur dalam	Umur dalam
8	82	2,4	46	1	1	Umur dalam	Umur dalam
9	90	2,5	90	1	2	Umur dalam	Umur sedang
10	76	2,6	65	1	0	Umur genjah	Umur genjah

Hasil dari Tabel 1 mengahsilakn.Pengujian Akurasi Keseluruhan Sistem Hasil dari pengujian akurasi keseluruhan sistem didapatkan presentase keberhasilan sebesar 80%. Hasil yang tertera pada dua pengujian yang dilakukan pada bab ini memperlihatkan bahwa kelas yang sering mengakibatkan kesalahan ialah umur sedang. Hal ini terjadi karena kelas umur sedang diapit oleh dua kelas lainnya atau bisa dikatakan berada di tengah dua kelas lainnya. Data dengan kelas yang berada ditengah akan memperbesar kemungkinan memiliki nilai fitur yang bersinggungan dengan nilai fitur milik kelas lain yang berujung pada kesalahan sistem. Kesimpulan yang didapatkan dari pengujian ini yaitu sistem menunjukan dapat melakukan klasifikasi masa panen dari jenis varietas unggul kedelai dengan cukup baik.

SIMPULAN DAN SARAN

KESIMPULAN

Kesimpulan hasil uji yang dilakukan pada pembahasan perancangan algoritma klasifikasi masa panen varietas unggul kedelai menggunakan K-Nearest Neighbor mendapatkan hasil yang cukup baik. Algoritma dirancang dengan menggunakan dua fungsi utama yaitu Preprocessing dan Klasifikasi KNearest Neighbor (KNN). proses dari preprocessing akan menggunakan normalisasi min-max untuk mendapatkan hasil rentang nilai dari data latih agar menjadi seragam antara satu fitur dengan fitur lainnya, sedangkan klasifikasi K-Nearest Neighbor (KNN), bertujuan untk mendapatkan nilai fitur dari data uji yang akan dihitung menggunakan euclian distance terhadap setiap data latih dengan menggunakan nilai K=5. Hasil pengujian pengaruh nilai K terhadap akurasi dengan menggunakan nilai K = 3,5,7. Pada pengujian nilai K= 3 mendapatkan akurasi 70%, pengujian nilai K=5 mendapatkan akurasi 80%, dan nilai K=7 mendapatkan hasil 50%. Hasil dari pengujian berikut bahwa nilai K=7 mendapatkan presentasi paling kecil

dikarenakan sistem menerima data yang lebih saat melakukan voting, sedangkan $K=5$ memiliki jumlah data yang cukup, tidak memiliki data sedikit ataupun lebih, sedangkan untuk $K=3$ memiliki kelas yang benar, tetapi masih memiliki sedikit kekurangan data untuk mencapai hasil yang optimal. 3. Hasil dari seluruh pengujian aplikasi keseluruhan sistem mendapatkan presentase keberhasilan sebesar 80%, dikarenakan kelas umur sedang diapit oleh dua kelas lainnya atau bisa dikatakan berada ditengah dua kelas lainnya. Hasil dari klasifikasi masa panen dari jenis varietas unggul kedelai mendapatkan hasil dari sistem dengan cukup baik.

SARAN

- a. Penerapan Metode k-Nearest Neighbors untuk Memprediksi Produktivitas Panen Kedelai Berdasarkan Data Cuaca dan Sifat Tanah
- b. Prediksi Hasil Panen Kedelai Menggunakan k-Nearest Neighbors dengan Fitur Meteorologi dan Karakteristik Tanah
- c. Model KNN untuk Peramalan Produktivitas Kedelai: Studi Kasus Menggunakan Data Iklim dan Parameter Tanah
- d. Penerapan k-Nearest Neighbors dan Seleksi Fitur untuk Prediksi Hasil Panen Kedelai Berbasis Data Cuaca-Tanah
- e. Pengembangan Sistem Prediksi Hasil Panen Kedelai Menggunakan k-Nearest Neighbors dan Optimasi Parameter K
- f. Perbandingan k-Nearest Neighbors dengan Algoritma Lain dalam Memprediksi Hasil Panen Kedelai Menggunakan Data Cuaca dan Tanah
- g. Integrasi Data Cuaca Historis dan Profil Tanah untuk Prediksi Hasil Panen Kedelai Menggunakan KNN
- h. Prediksi Produksi Kedelai dengan k-Nearest Neighbors: Pengaruh Normalisasi Data dan Penanganan Ketidakeimbangan
- i. Penerapan KNN pada Prediksi Hasil Panen Kedelai Berbasis Fitur Iklim Musiman dan Sifat Tanah Lokal
- j. Model Prediksi Hasil Panen Kedelai Menggunakan k-Nearest Neighbors dengan Ekstraksi Fitur dari Data Cuaca Harian

DAFTAR PUSTAKA

- Farokhah, L. (2020). Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(6), 1129. <https://doi.org/10.25126/jtiik.202072260>
- Hidayati, N., & Hermawan, A. (2021). KNearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation. *Journal of Engineering and Applied Technology*, 2(2), 86–91. <https://doi.org/10.21831/jeatech.v2i2.427>
- Irwan, A. W. (2006). Budidaya tanaman kedelai (*Glycine max* (L.) Merrill). Jurusan Budidaya Pertanian Fakultas Pertanian Universitas Padjadjaran, 1–43.
- Krithika, N., & Grace Selvarani, A. (2018). An individual grape leaf disease identificatio using leaf skeletons and KNN classification. *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICII ECS 2017*, 2018-Janua, 1–5. <https://doi.org/10.1109/ICII ECS.2017.82>
- Nur Mahdi, N., & Suharno, S. (2019). Analisis Faktor-Faktor Yang Memengaruhi Impor Kedelai Di Indonesia. *Forum Agribisnis*, 9(2), 160–184. <https://doi.org/10.29244/fagb.9.2.160-184>
- Prahudaya, T. Y., & Harjoko, A. (2017). Metode Klasifikasi Mutu Jambu Biji Menggunakan Knn Berdasarkan Fitur Warna Dan Tekstur. *Jurnal Teknosains*, 6(2), 113. <https://doi.org/10.22146/teknosains.2697>

- Puspitasari, D., Rahmad, C., & Astiningrum, M. (2016). Normalisasi Tabel Pada Basisdata Relasional. Jurnal Prosiding SENTIA | ISSN: 2085-2347, 8(1), 340–345.
- Sri Wahyuningsih, S. S. (2021). Analisis Kinerja Perdagangan Kedelai. Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian 2021.
- Supadi. (2016). *Impact of The Sustained Soybean Import on Food Security*. 87–102. <https://media.neliti.com/media/publications/56227-ID-dampak-impor-kedelaiberkelanjutan-terha.pdf>
- Yustanti, W. (2012). Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah. Jurnal Matematika Statistika Dan Komputasi, 9(1), 57–68.