Clusterisasi Varietas Benih Tanaman Padi Menggunakan Algoritma K-Means

Arwansyah, Suryani
STMIK Dipanegara Makassa
arwansyah@dipanegara.ac.id, a.surya.a.z@gmail.com

Abstrak

Tanaman padi adalah salah satu komuditi yang sangat penting bagi msyarakat di Indonesia hal ini di karenakan makanan pokok masyarakat adalah nasi yang berasal dari tanaman padi. Setiap wiilayah di Indonesia memiliki lahan pertanian yang sangat luas yang membudidayakan tanaman padi. Berbagai system irigasi pun dibangun untuk mendukung produktifitas tanaman pokok tersebut. Pertumbuhan tanaman padi harus di dukung oleh beberapa factor utama yang meliputi benih unggul, system irigasi, tanah yang subur, pupuk yang tepat, serta sinar matahari yang cukup. Pada penilitian ini akan diimplementasikan algoritma K-Mean untuk mengclustirasasi benih yang baik dan kurang baik berdasarkan beberapa factor yakni amylase level, rice rendement of milled, rice rendement of broken skin, dan texture. Hal ini sangat berguna sebab informasi yang dihasilkan dapat dijadikan langkah awal oleh petani dalam memilih benih berkualitas. Data yang digunakan berasal dari balai pertanian. Hasil dari penelitian ini menunjukkan bahwa algoritma K-Means dapat memproses data mengenai varietas benih dan menghasilkan informasi dalam bentuk cluster benih yang baik dan kurang baik.

Keywords : *Algoritma,K-Means,Varietas, Benih.*

Abstract

Rice plant is one of commodity that very important for society in Indonesia, because staple food our society is that rice which come from rice plant. Every region in Indonesia has agricultural land that very wide which cultivate rice plant. Various rice plant have to support by some main factor that cover superior seeds, irrigation system, fertile soil, right fertilizer, dan enough sunsine. In this research, will be apply K-Means algorithm in order to clustering good variety and bad variety according to several component that are amylase level, rice rendement of milled, rice rendement of broken skin and texture. It is extremely helpful because information that generate coule be first stage by farmers to choose superior seeds. Data that will be use came from agricultural bureau. The result of this research shows that K-Means algorithm can process data related variety of seeds and produce information in shape cluster that are good seeds and bad seeds.

Keywords*: Algorithm, K-Means, Sunsine, Rice Plant*

1. Pendahuluan

Teknologi informasi adalah teknologi yang dibangun dengan basis utama teknologi komputer. Perkembangan yang terus berlanjut dari teknologi membawa aplikasi utama teknologi ini pada proses pengolahan data yang berujung pada informasi. Teknologi informasi menjadi sebuah teknologi yang lebih luas pengaruh dan implikasinya dibandingkan teknologi komputer, yang awalnya hanya berkembang dalam dunia komputasi, hitung menghitung. Prinsipnya aplikasi teknologi informasi adalah alat bantu bagi Manusia untuk mengolah data menjadi informasi. Informasi ini kemudian dimanfaatkan oleh manusia, baik secara langsung maupun tidak langsung untuk menjalankan pekerjaannya. Penerapan teknologi informasi di dalam kehidupan akan selalu berkembang mengikuti kebutuhan manusia yang semakin kompleks dan bervariasi. Komponen dasar pembentuk teknologi selain teknologi komputer disebabkan karena berkembangnya bidang Telekomunikasi. Perkembangan telekomunikasi dianggap sebagai salah satu sebab utama munculnya revolusi informasi yang terjadi saat ini.

Tidak dapat disangkal bahwa salah satu penyebab utama terjadinya era globalisasi yang datangnya lebih cepat dari dugaan semua pihak adalah karena perkembangan pesat teknologi informasi.

Implementasi internet, electronic commerce, electronic data interchange, virtual office, telemedicine, intranet, dan lain sebagainya telah menerobos batas-batas fisik antar negara. Penggabungan antara teknologi komputer dengan telekomunikasi telah menghasilkan suatu revolusi di bidang sistem informasi. Data atau informasi yang pada jaman dahulu harus memakan waktu berhari-hari untuk diolah sebelum dikirimkan ke sisi lain di dunia, saat ini dapat dilakukan dalam hitungan detik. Tidak berlebihan jika salah satu pakar IBM menganalogikannya dengan perkembangan otomotif sebagai berikut: "seandainya dunia otomotif mengalami kemajuan sepesat teknologi informasi, saat ini telah dapat diproduksi sebuah mobil berbahan bakar solar, yang dapat dipacu hingga kecepatan maximum 10,000 km/jam, dengan harga beli hanya sekitar 1 dolar Amerika !". Secara mikro, ada hal cukup menarik untuk dipelajari, yaitu bagaimana evolusi perkembangan teknologi informasi yang ada secara signifikan mempengaruhi persaingan antara perusahaan-perusahaan di dunia, khususnya yang bergerak di bidang jasa.

Peningkatan ketahanan pangan di Indonesia harus dilakukan melalui kegiatan pembangunan yang berkelanjutan. Dengan Sistem pembangunan berkelanjutan dalam bidang pertanian, bangsa Indonesia suatu saat akan dapat mandiri dalam pemenuhan kebutuhan panggan. Peningkatan ketahanan pangan suatu bangsa saat ini tidak hanya dengan tindakan nyata dalam kegiatan pertanian, tetapi sudah menerapkan teknologi informasi pengolah data, Seperti Thailand, jepang, dan beberapa Negara di Asia. Pemanfaatan teknologi informasi dalam bidang pertanian telah diterapkan dapat digunakan untuk menganalisa kondisi seperti kondisi tanah, air, curah hujan, varietas atau benih, hingga kondisi pemasaran sehingga teknologi informasi merupakan hal penting yang harus di aplikasikan pada bidang pertanian.

2. Metode Penelitian

Dalam rangka keberhasilan penelitian, maka digunakan dua jenis metode penelitian untuk pengumpulan data yaitu :

1. Penelitian pustaka
   Penelitian dilakukan melalui buku-buku pustaka dan internet yang dapat memberikan teori-teori mengenai permasalahan yang diteliti, kemudian mencocokkan dengan kemungkinan-kemungkinan yang terjadi dalam usaha penyelesaian masalah.
2. Penelitian lapangan
   Penelitian yang dilakukan dengan mengunjungi langsung lokasi penelitian. Di tempat penelitian tersebut penulis melakukan pengamatan langsung terhadap objek penelitian dan melakukan tanya jawab kepada petani yang terkait.

2.1 Teknik Pengumpulan Data

Pengumpulan data adalah salah satu hal yang penting dilakukan dalam memperoleh data yang diinginkan. Data yang dikumpulkan tersebut akan menjadi sebuah basis data. Dengan adanya data yang diambil tersebut, akan sangat membantu sebagai bahan pertimbangan dalam analisis sistem. Adapun teknik yang digunakan dalam pengumpulan data yaitu :

1. Teknik Wawancara
   Teknik ini merupakan suatu teknik pengumpulan data dengan cara melakukan wawancara terhadap beberapa petani yang terkait yang berada di wilayah objek penelitian.
2. Teknik Observasi
   Teknik ini merupakan suatu teknik pengumpulan data dengan cara mengamati dan melihat langsung jenis – jenis varietas benih padi.

2.2 Alat dan Bahan Penelitian

1. Alat Penelitian:

a. Hardware
   1. 1 unit Notebook
   2. Processor inter celeron, ~2.0GHz
   3. Memory RAM DDR 2 GigaByte
   4. Harddisk 500 GB

b. Software
   1. Windows 10
   2. Microsoft Office Excel 2007

2.3 Tinjauan Pustaka

2.3.1 Data Mining

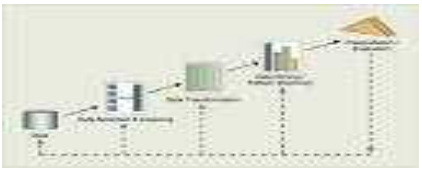
Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui. Data mining dapat diartikan sebagai analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, association rule, clustering, deskripsi dan visualisasi. Secara sederhana data mining bisa dikatakan sebagai proses menyaring atau menambang pengetahuan dari sejumlah data yang besar. [1].

Tujuan Dari Adanya Data Mining adalah:

1. Explanatory, yaitu untuk menjelaskan beberapa kegiatan observasi atau suatu kondisi.
2. Confirmatory, yaitu untuk mengkonfirmasi suatu hipotesis yang telah ada.
3. Exploratory, yaitu untuk menganalisa data baru suatu relasi yang janggal

Proses Data Mining

1. Pembersihan data (Data Cleaning), untuk membersihkan noise dan data yang tidak konsisten.
2. Integrasi Data, penggabungan data dari berbagai sumber.
3. Transformasi data, data diubah menjadi bentuk yang sesui untuk di mining.
4. Aplikasi teknik data mining, proses dimana teknik data mining diterapkan untuk mengekstrak pola-pola tertentu pada data.
5. Evaluasi pola yang ditentukan.
6. Presentasi pengetahuan, menggunakan teknik visualisasi untuk menampilkan hasil data mining kepada pengguna



Gambar 2.1 Tahapan dalam proses data mining [1]

Tools Data Mining

1. Klasifikasi
   Klasifikasi merupakan tools data mining yang paling umum. Ciri-ciri klasifikasi adalah memiliki definisi yang jelas tentang kelas-kelas dan training set. Klasifikasi bertujuan memprediksi kelas dari suatu data yang belum diketahui kelasnya. Dalam mencapai tujuannya tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu.
2. Estimasi
   Estimasi hamper sama dengan klasifikas namum estimasi lebih menangani data kontinyu. Contoh estimasi antara lain memperkirakan jumlah anak dalam keluarga, memperkirakan pendapatan keluarga, dan memperkirakan nilai probabilitas pemegang kartu kredit terhadap pada hal yang ditawarkan oleh pihak bank, misalnya tawaran untuk pemasangan iklan dengan tema olah raga ski pada amlop tagihan.
3. Prediksi
   Prediksi juga hamper sama seperti klasifikasi maupun estimasi, namun prediksi berusaha memprediksikan atau memperkirakan nilai atribut kelas dari suatu data untuk masa yang akan datang
4. Pengelompokan afinitas
   Pengelompokan afinitas adalah pengelompokan berdasarkan hal – hal yang cenderung dilakukan bersamaan. Misalnya pengelompokan barang – barang yang biasanya dibeli bersamaan dalam suatu supermarket.
5. Pengelompokan
   Pengelompokan adalah tugas data mining yang menggunakan metode membagi populasi yang heterogen menjadi sejumlah kelompok data yang homogeny. Pengelompokan tidak tergantung pada predefined classes dan training set. Data dikelompokan berdasarkan cirri-ciri yang sama. Pengelompokan sering dijadikan sebagai pendahuluan dalam pemodelan data mining.
6. Deskripsi
   Deskripsi merupakan tugas sekaligus tujuan dari data mining, yaitu berusaha mendeskripsikan suatu yang sedang terjadi atau terdapat dalam suatu basis data yang rumit. Teknik yang memberikan deskripsi yang jelas misalnya teknik market basket analysis. [2]

2.3.2 Algoritma K-Means

Data mining berkembang menjadi alat bantu untuk mencari pola-pola yang berharga dalam suatu basisdata yang sangat besar jumlahnya, sehingga tidak memungkinkan dicari secara manual. Beberapa teknik data mining dapat diklasifikasikan ke dalam kategori berikut, meliputi klasifikasi, *clustering*, penggalian kaidah asosiasi, analisa pola sekuensial, prediksi, visualisasi data dan lain sebagainya.

Teknik *clustering* adalah teknik yang digunakan untuk menangani data yang besar dengan banyak atribut ke dalam sejumlah kelompok kecil. *Clustering* dilakukan dengan terlebih dahulu menganalisis bagian kecil dari data untuk menentukan klaster. *Clustering* merupakan pengelompokkan *record*, observasi, atau kasus ke dalam kelas-kelas objek yang mirip. *Clustering* berbeda dengan klasifikasi dimana dalam *clustering* tidak terdapat variabel target. Salah satu algoritma *clustering* adalah *K-Means*. Clustering merupakan suatu teknik data mining yang membagi-bagikan data ke dalam beberapa kelompok (grup atau cluster atau segmen) yang tiap cluster dapat ditempati beberapa anggota bersama-sama. Setiap obyek dilewatan ke grup yang paling mirip dengannya. Ini menyerupai menyusunan binatang dan tumbuhan ke dalam keluarga – keluarga yang para anggotanya mempunyai kemiripan. Clustering tidak mensyaratkan pengetahuan sebelumnya dari grup yang dibentuk, juga dari para anggota yang harus mengikutinya.[3]

Algoritma K-Means diperkenalkan oleh J.B. MacQueen pada tahun 1976, salah satu algoritma clustering sangat umum yang mengelompokkan data sesuai dengan karakteristik atau ciri-ciri bersama yang serupa. Grup data ini dinamakan sebagai cluster. Data di dalam suatu cluster mempunyai ciri-ciri (atau fitur, karakteristik, atribut, properti) serupa dan tidak serupa dengan data pada cluster lain.

Beberapa tahap dari algoritma *K-Means* dapat dilihat pada algoritma berikut :

1. Pemilihan secara acak *K*
2. Inisialisasi *k* pusat klaster (centroid) secara random
3. Tempatkan setiap data atau objek ke klaster terdekat.

Euclidean distance

$$d_{Euclidean}(x,y) = \sqrt{\sum_{j=1}^{p}(x_j - y_j)^2}$$

Ket p : dimensi data
x :jarak item data
y :jarak item data
d : distance

4. Hitung kembali pusat klaster dengan keanggotaan klaster yang sekarang. Pusat klaster adalah rata-rata (mean) dari semua data atau objek dalam klaster tertentu.

Algoritma *K-Means* dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya klaster yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari klaster atau biasa disebut dengan *centroid*, yang memiliki mean atau "*means*". Hitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Klasifikasikan setiap data berdasarkan kedekatannya dengan *centroid*. Lakukan langkah tersebut hingga nilai *centroid* tidak berubah (stabil).[3]

Algoritma K-Means

Berikut ini adalah bagaimana algoritma K-Means mempartisi suatu dataset ke dalam cluster-cluster:

1. Algoritma menerima jumlah cluster untuk mengelompokkan data ke dalamnya, dan dataset yang akan dicluster sebagai nilai input.
2. Algoritma kemudian membuat sebanyak K cluster awal (K= jumlah cluster yang diperlukan) dari dataset, sekaligus memilih K record data secara acak (random) dari dataset. Sebagai contoh, jika terdapat 10,000 baris data di dalam dataset dan 3 cluster perlu dibentuk, maka K=3 cluster awal pertama akan dibuat dengan mengambil 3 record secara random dari dataset, sebagai cluster permulaan. Setiap cluster awal yang dibentuk tersebt mempunyai hanya satu record data.
3. Algoritma K-Means menghitung rata-rata aritmatika dari setap cluster yang dibentuk dalam dataset. Rata-rata dari suatu cluster adalah rata-rata dari semua record yang terdapat di dalam cluster tersebut.

Karena di dalam semua K cluster pertama hanya ada satu record maka rata-ratanya adalah rata-rata record tersebut. Rata-rata dari suatu record adalah kumpulan nilai-nilai yang membangun record tersebut. Sebagai contoh jika di dalam dataset S terdapat record P yang menerima nilai-nilai untuk field Tinggi, Berat dan Usia, maka dapat ditulis P = {Usia, Tinggi, Berat). Jika John mempunyai Usia John = 20 tahun, Tinggi = 1.70 meter dan Berat = 80 Pon maka ditulis John = {20, 170, 80}. Karena terdapat hanya satu record di dalam setiap cluster awal maka rata-rata dari cluster dimana John berada adalah = {20, 170, 80}.

4. Selanjutnya, K-Means mengirim K record di dalam dataset ke hanya salah satu dari cluster awal. Record ke-4 sampai ke-6 dilewatkan ke cluster terdekat (nearest cluster, yaitu cluster yang sangat mirip dengan record tersebut) menggunakan suatu ukuran jarak atau kemiripan seperti ukuran jarak Euclidean atau Manhattan/City-Block.
5. K-Means mengkalkulasi ulang rata-rata aritmatika dari semua cluster. Rata-rata dari suatu cluster adalah rata-rata dari semua record di dalam cluster tersebut. Sebagai contoh, jika suatu cluster mengandung dua record John = {20, 170, 80} dan Henry = {30, 160, 120}, maka rata-rata P(rata-rata) dinyatakan sebagai P(rata-rata) = {Usia(rata-rata), Tinggi(rata-rata), Berat(rata-rata)). Usia(rata-rata) = (20 + 30)/2, Tinggi(rata-rata) = (170 + 160)/2 dan Berat(rata-rata) = (80 + 120)/2. Rata-rata aritmatik dari cluster ini adalah {25, 165, 100}. Rata-rata tersebut menjadi pusat dari cluster baru ini. Mengikuti prosedur yang sama, pusat-pusat cluster baru dibentuk untuk semua cluster yang telah ada.
6. K-Means mengirimkan lagi setiap record di dalam dataset ke hanya salah satu dari cluster-cluster baru yang terbentuk. Suatu record atau titik-titik data dilewatkan ke cluster terdekat, seperti sebelumnya.
7. Langkah-langkah sebelumnya diulang sampai terbentuk cluster-cluster stabil dan prosedur K-Means selesai. Cluster stabil terbentuk ketika iterasi atau perulangan dari K-Means tidak membuat cluster baru sebagai pusat cluster atau nilai rata-rata aritmatika dari semua cluster baru sama dengan cluster lama. Terdapat beberapa teknik untuk menentukan kapan suatu cluster stabil terbentuk atau kapan algoritma K-Means berakhir. [4]

Definisi matematis jarak Euclidean dan Manhattan

Ukuran Jarak Euclidean antara dua titik atau obyek atau item X dan Y didefinisikan oleh persamaan: Jarak Euclidean (X,Y) = (|X1-Y1|2 + |X2-Y2|2 + … + |XN-1-YN-1|2 + |XN-YN|2)1/2 Sedangkan ukuran jarak Manhattan atau City Block didefinisikan sebagai berikut: Jarak Manhattan (X,Y) = (|X1-Y1| + |X2-Y2| + … + |XN-1-YN-1| + |XN-YN|). [3]

Ciri-Ciri K-means yaitu:

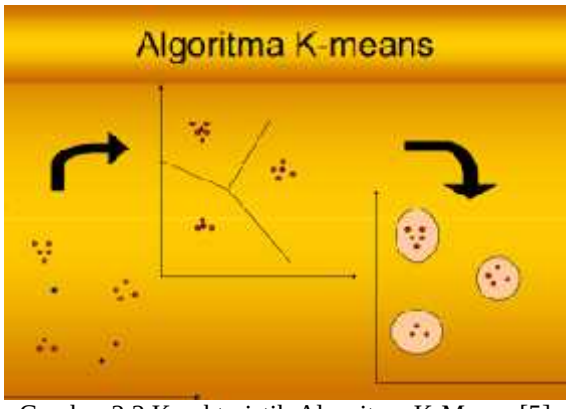
1. K- Means algoritma sangat terkenal karena kemudahan dan kemampuannya untuk mengklaster data besar dan data outliet dengan sangat cepat
2. Setiap data harus masuk ke cluster tertentu
3. Data masuk ke cluster tertentu dan dapat berpindah ke cluster yang lainnya pada tahap proses berikutnya. [4]

Langkah Langkah Algoritma K-Means :

1. Tentukan k sebagai jumlah cluster yang ingin dibentuk
2. Bangkitkan k centeroid awal secara random
3. Hitung jarak setiap data ke masing masing centeroid
4. Setiap data memilih centeroid yang terdekat
5. Tentukan posisi centeroid baru dengan cara menghitung nilai rata rata dari data data yang memilih pada centeroid yang sama
6. Kembali ke langkah 3 jika posisi centeroid baru dengan centeroid lama tidak sama.[5]

Karateristik K-Means

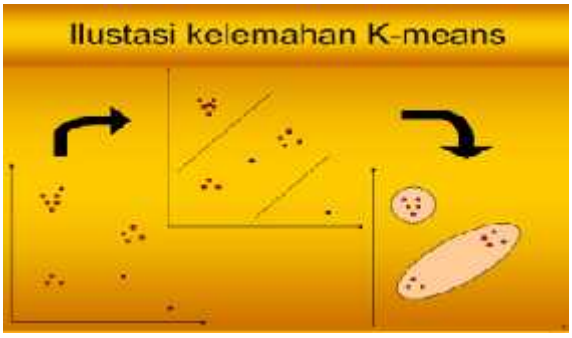
1. K-Means dangat cepat dalam proses clustering
2. K-Means sangat sensitive pada pemangkitan centeroid awal secara random
3. Memungkinkan suatu cluster tidak mempunyai anggota
4. Hasil clusteringdengan k means bersifat tidak unik (slalu berubah ubah) terkadang baik terkadang jelek
5. K-Means sangat sulit untuk mencapai global optimum.[5]



Gambar 2.2 Karakteristik Algoritma K-Means.[5]

Kelemahan Algoritma K-Means

K-Means merupakan metode klasterisasi yang paling terkenal dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengklaster data yang besar, mampu menangani data outlier, dan kompleksitas waktunya linear O(nKT) dengan n adalah jumlah dokumen,K adalah jumlah kluster, dan T adalah jumlah iterasi. K-means merupakan metode pengklasteran secara partitioning yang memisahkan data ke dalam kelompok yang berbeda. Dengan *partitioning* secara iteratif, KMeans mampu meminimalkan rata-rata jarak setiap data ke klasternya.
Metode ini dikembangkan oleh Mac Queen pada tahun 1967. [5]



Gambar 2.3 Kelemahan Algoritma K-Means.[5]

Clustering

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan(kemiripan).Clustering beda dengan group, kalau group berarti kelompok yang sama,kondisinya kalau tidak ya pasti bukan kelompoknya.Tetapi kalau cluster tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sample yang ada, salah satunya dengan menggunakan rumus jarak ecluidean.Aplikasinya cluster ini sangat banyak, karena hamper dalam mengidentifikasi permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja. [6]

Manfaat Clustering

1. Identifikasi obyek (Recognition) : Dalam bidang mage Processing , Computer Vision atau robot vision
2. Decission Support System dan data mining. Segmentasi pasar, pemetaan wilayah, Manajemen marketing dll. [6]

Prinsip Dasar Clustering

1. Similarity Measures (ukuran kedekatan)
2. Distances dan Similarity Coeficients untuk beberapa sepasang dari item Ecluidean Distance:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_p - y_p)^2}$$

Atau :

$$d(x,y) = \left[\sum_{i=1}^{p} | x_i - y_i |^2 \right]^{1/2}$$

Algoritma Clustering

1. Partisi item menjadi K initial cluster
2. Lakukan proses perhitungan dari daftar item, tandai item untuk kelompok yang mana berdasarkan pusat(mean) yang terdekat (dengan menggunakan distance dapat digunakan Euclidean distance).Hitung kembali pusat centroid untuk item baru yang diterima pada cluster tersebut dari cluster yang kehilangan item.
3. Ulangi step 2 hingga tidak ada lagi tempat yang akan ditandai sebagai cluster baru. [7]

2.3.3 Varietas Padi

Padi adalah salah satu makanan pokok paling banyak dikonsumsi di seluruh dunia. Makanan ini dikonsumsi terutama di Asia dan Amerika Selatan.Padi, dengan nama ilmiah *Oryza sativa* L. adalah tanaman yang dibudidaya, meski ada juga yang merupakan padi liar. Padi sendiri diduga dimulai dari India atau Indocina, namun dibudidayakan di Indonesia sekitar 1500 SM. Di negara agraris seperti Cina, India, Bangladesh, dan Indonesia, padi merupakan tanaman utama. Padi jadi penghasil sebagian besar makanan pokok konsumsi masyarakat.

Jenis-jenis padi berdasarkan varietas

1. Varietas hibrida

   Varietas hibrida adalah varietas padi yang hanya sekali tanam. Kelebihan padi varietas hibrida adalah potensi hasil panen yang maksimal. Hasil panen dapat mencapai dua kali lipat dari padi lokal. Butiran padi yang dihasilkan lebih bagus, dengan kualitas nasi yang lebih pulen dan wangi.

   Namun varietas hibrida sendiri memiliki kelemahan, yaitu kualitas hasilnya akan berkurang jauh apabila berasal dari tanaman turunannya. Artinya, padi harus berasal dari bibit original, karena apabila hasil panen kemudian ditanam ulang, hasil ini akan berbeda dengan bibit aslinya. Harga benih varietas hibrida ini termasuk yang termahal.

   Jenis varietas padi hibrida antara lain Intani 1 dan 2, Rokan, SL 8 dan 11 SHS, Segera Anak, PP1, H1, Bernas Prima, SEMBADA B3, B5, B8 DAN B9, Long Ping (pusaka 1 dan 2), Adirasa-1, Adirasa-64, Hibrindo R-1, Hibrindo R-2, Manis-4 dan 5, Hipa4, Hipa 5 Ceva, Hipa 6 Jete, Hipa 7-10 11, MIKI 1-3, SL 8 SHS, SL 11 HSS dan Maro.
2. Varietas padi unggul

   Varietas padi unggul berada satu tingkat di bawah varietas hibrida. Varietas ini dapat ditanam berkali-kali dengan kualitas yang sama. Artinya, hasil panen dari varietas padi unggul masih bisa dijadikan benih. Harga benih padi unggul pun tidak semahal benih padi hibrida. Untuk hasil produksi pun padi unggul dapat dikatakan baik, dapat mencapai 8-10 ton per hektar. Beberapa contoh varietas padi unggul antara lain adalah Inpara 1-8, Inpago 1-5, Inpari 1-21, Inpari 31, Inpari 33, Inpari 34 Salin Agritan, dan Inpari 35 Salin Agritan. Varietas padi unggul pun ada juga yang dikembangkan dan dirilis oleh pemerintah, seperti Inpari 34 dan Inpari 35. Keunggulan varietas ini adalah ketahanannya terhadap hama wereng cokelat.
3. Varietas padi local

   Varietas padi lokal adalah varietas padi yang khusus berada di daerah tertentu. Varietas semacam ini hanya cocok ditanam di daerah tertentu saja, karena membutuhkan spesifikasi khusus untuk tumbuh dan memproduksi padi. Padi lokal biasanya menghasilkan produksi sekitar 7-8 ton per hektar. Rasa beras dari padi lokal juga kurang enak. Jenis jenis padi lokal antara lain : Gropak (Kulon Progo), Indramayu, Dharma Ayu, Srimulih, Andel Jaran, Merong, Gundelan, Marong, Simenep, dan Ketan Lusi.

Jenis-jenis padi berdasarkan beras yang dihasilkan

1. Padi ketan

   Padi ketan lebih lengket dari padi nasi, sehingga tidak dijadikan makanan pokok. Padi ketan biasanya dijadikan bahan pembuatan tape ketan, bubur ketan, dan macam-macam makanan khas daerah. Varian padi ketan antara lain adalah varian padi ketan merah, putih, dan hitam.
2. Padi wangi

Sesuai namanya, padi wangi memiliki karakteristik beraroma wangi. Padi seperti ini contohnya adalah padi pandanwangi.

3. Padi pera
   Padi pera adalah padi yang apabila berasnya dimasak, akan menghasilkan nasi bertekstur pera. Pera adalah tekstur nasi yang sedikit keras. Tekstur ini berasal dari kadar amilosa yang tinggi. Semakin tinggi kadar amilosa, semakin terasa tekstur nasi tersebut. Kadar amilosa yang menghasilkan tesktur pera minimal 25%. Padi pera diproduksi dan populer di daerah Sumatera Barat dan Riau. Padi dengan kadar amilosa tinggi tak hanya dijadikan nasi, pun juga menjadi bahan utama pembuatan bihun dan tepung beras. Contoh padi pera adalah Inpari 12 (amilosa 26,4%), Inpara 1 (27,9%), Inpara 3, (28,6%), Inpara 4 (29%), Inpari 17 (amilosa 26%), dan Hipa 4 (24,7%).
4. Padi pulen
   Padi pulen adalah padi yang apabila berasnya dimasak, akan menghasilkan karakteristik nasi yang pulen. Sebagian orang lebih menyukai nasi yang pulen alias sedikit lengket. Pulen berasal dari amilopektin yang tinggi di dalam padi dan kadar amilosa di bawah 25%. Apabila dimasak, nasi yang dihasilkan akan berasa sedikit lengket. Contoh padi dengan tekstur yang pulen adalah tekstur padi Inpari 13, Ciherang, dan IR64.

Jenis-jenis padi berdasarkan budidaya

1. Padi gogo
   Padi gogo adalah jenis padi yang tidak ditanam di sawah seperti pada umumnya. Jenis padi ini ditanam di kebun atau di ladang. Kelebihan padi gogo adalah tidak memerlukan irigasi khusus. Daerah yang sering mengembangkan padi gogo adalah daerah tadah hujan, contohnya di Lombok. Di Lombok sendiri ada sistem yang disebut padi gogo rancah. Ia memberikan penggenangan di selang waktu tertentu saja, sehingga hasil padinya meningkat. Pada lahan-lahan yang mengembangkan padi gogo, biasanya diterapkan juga sistem bercocok tanam tumpang sari. Artinya, petani tidak hanya menanam padi, namun juga disandingkan dengan tanaman produksi lainnya, seperti jagung atau ketela/singkong.
2. Padi rawa
   Padi rawa adalah padi yang sering ditanam di persawahan. Padi ini membutuhkan genangan air, sehingga perlu diirigasi secara konsisten.

Jenis-jenis padi berdasarkan kelas benih

1. Benih penjenis
   Benih penjenis merupakan benih yang diproduksi dan diawasi langsung oleh Pemulia Tanaman atau instansi terkait. Benih ini disesuaikan dengan karakteristik lahan tertentu, dan dikembangkan agar tahan terhadap hama yang ada di daerah tersebut.
2. Benih dasar
   Benih dasar adalah keturunan pertama dari Benih Penjenis. Benih dasar ini diproduksi di Instansi yang ditunjuk oleh Dirjen Tanaman Pangan, denga bimbingan dan pengawasan yang ketat. Untuk dapat diproduksi, benih ini perlu mendapatkan sertifikasi oleh Balai Pengawasan dan Sertifikasi Benih.
3. Benih Pokok (BP)
   Benih Pokok (BP) adalah keturunan dari Benih Dasar yang diproduksi dan dipelihara sedemikian rupa sehingga indetitas dan tingkat kemurnian varietas yang ditetapkan dapat dipelihara dan memenuhi standart mutu yang di tetapkan dan harus disertifikasi sebagai Benih Pokok oleh Balai Pengawasan dan Sertifikasi Benih.
4. Benih Sebar (BR)
   Benih Sebar (BR) adalah keturunan dari Benih Pokok. Benih sebar diproduksi dan dipelihara sedemikian rupa, dengan tujuan agar identitas dan kemurnian varietas bisa dipelihara dengan baik, memenuhi standar mutu yang ditetapkan, dan juga harus disertifikasi sebagai Benih Sebar oleh Balai Pengawasan dan Sertifikasi Benih.

3. Hasil dan Pembahasan

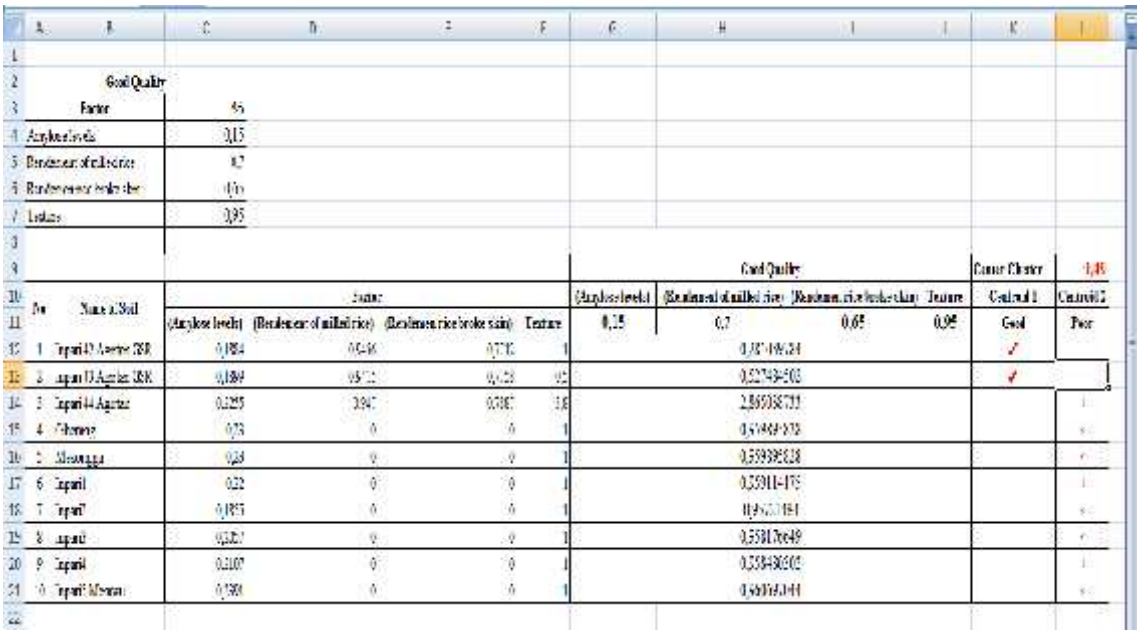
Ada beberapa tahap yang dilakukan dalam analisis algoritma K-Means yang meliputi :

1. Menentukan mengumpulkan informasi mengenai data varietas padi
2. Mengumpulkan informasi mengenai ciri-ciri varietas padi yang baik
3. Menghitung jarak setiap data dengan mengambil nilai rata-rata dari ciri-ciri varietas padi yang baik

4. Mengelompokkan jarak data terdekat dengan centroid

Tabel hasil analisa algortima K-NN dapat dilihat pada tabel 3.1 berikut :

Table 3.1 Hasil Perhitungan Algoritma K-NN



4. Kesimpulan

Berdasarkan hasil pembahasan menggunakan algoritma K-Means maka dihasilkan kesimpulan :

1. Memilih varietas padi yang baik yang sesuai dengan kondisi atau factor penentu keberhasilan produksi tanaman padi perlu system atau metode sehingga produksi panen akan menjadi lebih meningkat.
2. Bidang pertanian dapat memanfaatkan teknologi computer dalam memberikan informasi yang relevan dan berguna kepada petani atau unsut terkait sehingga hasil pertanian terutama hasil pangan dapat lebih meningkat dan berkualitas
3. Data mining dapat dimanfaatkan dalam bidang pertanian untuk menghasilkan informasi dan pengetahuan kepada petani dalam mengelola budidaya tanaman mereka
4. Algoritma K-Means dapat menghasilkan informasi berupa varietas benih padi yang berkualitas untuk setiap musim tanam.

Daftar Pustaka

[1] Kusrini, & Emha Taufik Luthfi. (2009). Algoritma Data mining. Yogyakarta: Andi.

[2] Fayyad, Usama. 1996. Advances in Knowledge Discovery and Data Mining. MIT Press.

[3] Agusta, Y. 2007. *K-means - Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol. 3 (Februari 2007): 47-60.

[4] Wieta B. Komalasari.2007. Metode Pohon Regresi Untuk Eksploratori Data Dengan Peubah Yang Banyak Dan Kompleks. Jurnal Informatika Pertanian Vol 16 No.1, Juli 2007

[5] Azis, Anifuddin, Sunarminto, Hendro., Medhanita, Dewi Renanti (2006). *Evaluasi Kesesuaian Lahan Untuk Budidaya Tanaman Pangan Menggunakan Jaringan Syaraf Tiruan*

[6] Witten, I. H., Frank, E., Hall, M. A. 2011. Data Mining Practical Machine Learning Tools and Techniques (3rd ed). USA: Elsevier

[7] Sevani,Nina.,Marimin. Sukoco,Heru 2009, "Sistem Pakar Penentuan Kesesuaian Lahan Berdasarkan Faktor Penghambat Terbesar (Maximun Limitian Factor) Untuk Tanaman Pangan", Jurnal Informatika Vol.10 No 1.

[8] Balai Penelitian Tanah. 2005.Analisis Kimia Tanah, Tanaman, Air, Dan Pupuk. Badan Penelitian dan Pengembangan Pertanian Depertemen Pertanian. Pulung, M. A. 2009. Kesuburan tanah. Penerbit Universitas Lampung. Bandar Lampung. 45 p.

.