BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Dalam penelitian ini, penulis menggunakan beberapa sumber pustaka. Sumber pustaka yang dimasukkan untuk digunakan sebagai pedoman dan pembandingan dalam penelitian yang akan penulis lakukan. Pustaka yang digunakan ditinjau dari segi kasus penelitian, objek penelitian dan metode yang digunakan dalam penelitian.

Penelitian yang pernah dilakukan oleh Felicia (2014) menerangkan bahwa sebuah gambaran yang menunjukkan pengelompokan daerah berdasarkan hasil pertanian padi dengan menggunakan metode *K-Means*.

Peneliti yang pernah dilakukan oleh Ody Adrian (2016) yang menggunakan metode KNN menerangkan bahwa aplikasi ini dapat memberikan rekomendasi pemilihan laptop secara baik dalam mengimplementasikan metode KNN dari perhitungan manual dan hasil dari aplikasi. Peneliti yang pernah dilakukan oleh Ruslan Abdul Gani (2016) dengan menggunakan metode KNN menerangkan bahwa aplikasi diagnosa tingkat stres dengan metode KNN dapat digunakan untuk mendiagnosa tingkat stres seseorang pengguna serta memberikan saran-saran sesuai tingkat stress yang dialami.

Rahayu dkk, (2014) menyatakan bahwa teknik klaster untuk mengelompokkan data yang dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Metode *K-Means* yang

digunakan pada penelitian untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok. Hasil yang didapat dalam penelitian ini adalah pengelompokan potensi kejahatan daerah dikota Banjarbaru diproses berdasarkan peajaran.

Table 2.1 Perbandingan Hasil Penelitian

Penulis dan Tahun	Objek	Metode	Hasil
Liana Felicia (2014)	Tanaman Padi	Algoritma <i>K-Means</i>	Sebuah gambaran yang menunjukkan pengelompokan daerah berdasarkan hasil pertanian padi.
Ody Andrian (2016)	Laptop	K-Nearest Neighbor	Aplikasi ini dapat memberikan rekomendasi pemilihan laptop secara baik dalam mengimplementasikan metode KNN dari perhitungan manual dan hasil dari aplikasi
Ruslan Abdul Gani (2016)	Manusia	K-Nearest Neighbor	Aplikasi diagnosa tingkat stres dengan metode KNN dapat digunakan untuk mendiagnosa tingkat stres seseorang pengguna serta memberikan saran-saran sesuai tingkat stress yang dialami.
Sri Rahayu, dkk (2014)	Kendaraan Bermotor	<i>K-Means</i>	Pengelompokan potensi kejahatan daerah dikota Banjarbaru diproses berdasarkan peajaran.
Vika Indiya Sari (2017)	Beras	<i>K-Means</i> dan <i>K-Nearest Neighbor</i>	Aplikasi diharapkan mampu membantu membantu masyarakat mempermudah proses klasifikasi mutu giling beras dengan menggunakan konsep data mining dan menghasilkan kualitas yang baik.

Pada penelitian yang akan penulis lakukan, diharapkan pencapaian hasil untuk klasifikasi kualitas beras mendapatkan kualitas baik dengan metode *K-Means* dan metode KNN bisa mencapai angka 90%.

2.2. Dasar Teori

2.2.1. Data Mining

Data mining adalah campuran dari statistik, kecerdasan buatan dan riset basis data yang masih berkembang (Gonunesco, 2011). Data mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama dan belum melewati masa ‘remaja’, maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya.

Data mining merupakan bidang yang kurang populer di awal kemunculannya. Namun di era komputerisasi ini dimana jumlah data berkembang sangat pesat, data mining mendapatkan popularitas yang sangat besar karena dapat diaplikasikan dalam berbagai jenis data dan berbagai bidang ilmu, misalnya: ilmu pendidikan, ilmu pemerintahan dan ilmu kesehatan. Pada umumnya, data mining muncul dari banyaknya jumlah data yang tersimpan dalam suatu basis data. Sehingga dari banyaknya data yang dimiliki dapat digali untuk memperoleh suatu pengetahuan yang bermanfaat.

2.2.2. *K-Means*

Algoritma *K-Means* merupakan algoritma pengelompokan iterative yang melakukan partisi set data ke dalam sejumlah K kluster yang sudah ditetapkan di

awal. Algoritma *K-Means* sederhana untuk diimplementasikan dan dijalankan, relative cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-Means* menjadi salah satu algoritma yang paling penting dalam bidang data mining (wu dan Kumar, 2009).

Parameter yang harus dimasukkan ketika menggunakan algoritma *K-Means* adalah nilai K. Nilai K yang digunakan biasanya didasarkan pada informasi yang diketahui sebelumnya tentang berapa banyak kluster data yang muncul dalam X, berapa banyak kluster yang dibutuhkan untuk penerapannya, atau jenis kluster yang dicari dengan mengeksplorasi/melakukan percobaan dengan beberapa nilai K.

Algoritma ini memiliki syarat yaitu nilai K harus sudah diketahui sebelumnya. Berikut merupakan tahapan-tahapan dalam algoritma *K-Means*:

1. Menentukan nilai k, yaitu jumlah kluster yang akan dibentuk.
2. Tentukan *centroid* (titik pusat kluster) sebanyak k dari data secara acak.
3. Hitung jarak setiap tiap objek data ke masing-masing *centroid* dengan menggunakan rumus *Euclidean distance*.
4. Kelompokkan tiap data berdasarkan jarak terdekat dengan *centroid*.
5. Tentukan *centroid* baru missal dengan cara menghitung nilai mean dari tiap kluster yang telah dibentuk.
6. Ulangi langkah 3-5 sampai tidak terjadi perubahan lagi.

2.2.3. Konsep Nearest Neighbor

Nearest Neighbor (NN) menjadi salah satu metode dalam 10 top metode data mining yang paling populer digunakan (Wu dan Kumar, 2009). Metode NN murni termasuk dalam klasifikasi yang *lazy learner* karena menunda proses pelatihan sampai ada data uji yang ingin diketahui label kelasnya, maka metode baru akan menjalankan algoritmanya. Algoritma NN melakukan klasifikasi berdasarkan kemiripan suatu data dengan data yang lain (Tan *et al*, 2005).

2.2.4. K-Nearest Neighbor

Metode KNN menjadi salah satu metode berbasis NN yang paling tua dan populer. Nilai K yang digunakan disini menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas pada data uji. Dari K tetangga terdekat yang terpilih kemudian dilakukan *voting* kelas dari K tetangga terdekat tersebut. Kelas dengan jumlah suara tetangga terbanyaklah yang diberikan sebagai label kelas hasil prediksi pada data uji tersebut (Tan *et al*, 2005).

Berikut merupakan tahapan-tahapan dalam algoritma KNN :

1. Menentukan parameter K = jumlah tetangga terdekat.
2. Hitung jarak antara data yang akan dievaluasi dengan semua data uji menggunakan rumus *Euclidian* seperti ditunjukkan pada persamaan 2.1 berikut.

$$d_{i=1}^p = \sqrt{\sum_{i=1}^p (X_{2i} - X_{1i})^2} \dots\dots\dots (2.1)$$

Keterangan :

X1 = data latih

X_2	= data uji
i	= variable data
d	= jarak data
p	= dimensi data

- Urutkan jarak data latih secara *ascending*.
- Cari jumlah kelas dari tetangga yang terdekat, dan tetapkan kelas tersebut sebagai kelas tersebut sebagai kelas yang akan dievaluasi.

2.2.5. Euclidean Distance

Euclidean Distance merupakan metrika yang sering digunakan untuk menghitung kesamaan 2 vektor. *Euclidean distance* menghitung akar dari kuadrat perbedaan 2 vektor (root of square differences between 2 vectors). Rumus dari Euclidean Distance pada persamaan 2.2 :

$$D(x_2, x_1) = \|x_2 - x_1\| = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \dots\dots\dots (2.2)$$

2.2.6. Beras

Beras merupakan hasil utama yang diperoleh dari proses penggilingan gabah hasil tanaman padi yang seluruh lapisan sekamnya terkelupas dan seluruh bagian lapisan bekatulnya telah dipisahkan.

Parameter input mutu giling beras yaitu :

1. Derajat Sosoh

Derajat sosoh yaitu tingkat terlepasnya lapisan bekatul dan lembaga dari butir beras. derajat sosoh dibagi menjadi 3 yaitu sebagai berikut :

- a. Derajat sosoh 100 % : tingkat terlapsnya seluruh lapisan bekatul, lembaga dan sedikit endosperm dari butir beras.
- b. Derajat sosoh 95% : tingkat terlepasnya sbagian besar lapisan bekatul, lembaga dan sedikit endosperm dari butir bras sehingga sisa yang belum terlepas sebesar 5%.
- c. Derajat sosoh 85% : tingkat terlepasnya sbagian besar lapisan bekatul, lembaga dan sedikit endosperm dari butir bras sehingga sisa yang belum terlepas sebesar 15%.

2. Kadar Air

Kadar air yaitu jumlah kandungan air di dalam butir beras yang dinyatakan dalam satuan butir berat basah.

3. Menir

Beras menir yaitu butir beras baik sehat maupun cacat yang mempunyai ukuran kecil dari 0,25 bagian butir beras utuh.

4. Butir Kapur

Butir kapur yaitu butir beras yang separuh bagian atau lebih berwarna putih seperti kapur dan bertekstur lunak yang disebabkan faktor fisiologi.

5. Butir kuning rusak

Butir kuning yaitu butir beras utuh, beras kepala, beras patah dan menir yang berwarna kuning atau kuning.

Cara menentukan presentase mutu giling beras yaitu sebagai berikut :

- a. Menentukan drajat sosoh dilakukan pada beras secara visual dengan indra penglihatan menggunakan pertolongankaca pembesar yang dibandingkan.

Pengukuran derajat sosoh sebagai berikut :

1. Timbang sampel beras yang akan dianalisis sebanyak kurang lebih 100 gram.
2. Pengukuran dilakukan secara visual dengan pertolongan kaca pembesar dan dibandingkan dengan contoh pembanding (standar) yang mempunyai derajat sosoh 100%, 90% dan 80%.

- b. Menentukan kadar air beras dilakukan dengan “*Air Oven Method*” atau dengan moisture tester elektronik yang telah dikalibrasi dengan standar oven. Penetapan kadar air metode oven adalah sebagai berikut :

1. Sampel beras sbanyak 5 gram ditimbang dalam cawan yang telah diketahui berat tetapnya.
2. Kemudian dikeringkan dalam cawan oven pada suhu 105 drajat C selama 3 jam atau sampai berat tetap.
3. Disimpan dalam desikator setelah dingin ditimbang.
4. Kadar air beras dihitung sebagai % fraksi massa

$$\text{Kadar air beras} = \frac{B-C}{B-A} \times 100\% \dots\dots\dots(2.3)$$

Keterangan :

A : berat cawan

B : berat contoh + cawan

C : berat contoh kering + cawan

Bb : berat basah

c. Menentukan butir menir dilakukan dengan cara pemisahan beras menggunakan ayakan berdiameter 2,0 mm. Cara menentukannya sebagai berikut :

1. Timbang 100 gram sampel beras.
2. Kemudian dipisahkan antara beras kepala / butir menir dengan menggunakan alat teste *Rice Grader*. Butir patah / menir dipisahkan dengan menggunakan ayakan berdiameter 2,0 mm atau menggunakan pinset dan kaca pembesar secara visual.
3. Timbang bobot butir menir.

$$\text{Presentase butir menir} = \frac{\text{berat butir menir}}{\text{berat contoh beras}} \times 100\% \dots\dots\dots(2.4)$$

d. Menentukan butir kuning rusak dilakukan dengan cara pemisahan beras menggunakan ayakan berdiameter 2,0 mm. Cara menentukannya sebagai berikut :

1. Timbang 100 gram beras.
2. Kemudian dipisahkan secara visual menggunakan pinset dan kaca pembesar.
3. Timbang bobot butir kuning/rusak.

$$\text{Kadar butir kuning / rusak} = \frac{\text{berat butir kuningrusak}}{\text{berat beras}} \times 100\% \dots\dots(2.5)$$

e. Menentukan butir kapur dilakukan dengan cara pemisahan beras menggunakan ayakan berdiameter 2,0 mm. Cara menentukannya sebagai berikut :

1. Timbang 100 gram beras.

2. Kemudian dipisahkan secara visual menggunakan pinset dan kaca pembesar.
3. Timbang bobot butir kapur.

$$\text{Kadar butir kapur} = \frac{\text{berat butir kapur}}{\text{berat beras}} \times 100\% \dots\dots\dots(2.6)$$