

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Jagung adalah tanaman pangan penting kedua di Indonesia setelah padi. Pertama kali dikenalkan sejak abad 15 oleh bangsa Portugis. Tanaman ini berasal dari benua Amerika yang telah lama dikenal dan dibudidayakan sejak ribuan tahun silam oleh manusia. Di Indonesia, jagung tersebar di berbagai wilayah yaitu Sumatera Utara, Sumatera Selatan, Lampung, Jawa Tengah dan Jawa Timur, Nusa Tenggara, Sulawesi Utara dan Selatan sampai Maluku. Daerah Jawa Timur merupakan 8 produsen utama jagung, sekitar 40% dari hasil nasional. Produksi jagung secara nasional, selama lima tahun terakhir rata-rata mencapai 9.740.600 ton, dengan lahan 3.750.000 ha dengan kenaikan 5,1%. Meskipun demikian, Indonesia masih mengimpor jagung rata-rata 1-2 juta ton/tahun karena kebutuhan jagung terutama untuk bahan baku pakan ternak terus meningkat. (Rohman, 2014)

Pada lima tahun terakhir, yaitu dari tahun 2013-2017 produksi jagung nasional terus mengalami peningkatan. Pada tahun 2013 Indonesia mampu memproduksi jagung sebesar 18,51 juta ton dan pada tahun 2017 produksinya sudah mencapai 27,95 juta ton. Daerah yang paling banyak menyumbangkan produksi jagung adalah Jawa Timur, yaitu sebesar 6,18 juta ton pada tahun 2017. Peringkat selanjutnya adalah daerah Jawa Tengah yang menyumbangkan produksi sebesar 3,51 juta ton. Secara keseluruhan produksi jagung tahun 2017 meningkat 18,55% dari tahun sebelumnya (Yonida, 2018)

Jagung merupakan tanaman yang memiliki nilai ekonomi tinggi, oleh karena itu dilakukan salah satu cara yang digunakan untuk meningkatkan produksi jagung yaitu dengan menerapkan teknik *Nested Association Mapping* (NAM) yang dirancang untuk mengidentifikasi dan membedah arsitektur genetik ciri-ciri kompleks dalam jagung (*Zea mays*). Teknik NAM memiliki kelebihan yaitu *scan*

genom hemat biaya daya tinggi melalui simulasi komputer berdasarkan data penanda empiris dan sifat simulasi dengan kompleksitas yang berbeda. Hasil teknik *NAM* ini berisi informasi tentang nilai ekspresi gen dari berbagai jenis variasi jagung berdasarkan organ dan nenek moyang tanaman tersebut. Ekspresi gen adalah proses dimana informasi yang terkandung dalam gen menjadi produk yang bermanfaat. Data ekspresi gen mentah dalam jumlah yang besar menyebabkan tantangan statistik dan analisis dalam memasukkan klasifikasi dataset menjadi kelas yang sebenarnya. Kendala pada analisis ekspresi gen adalah variabel pada data ekspresi gen sangatlah banyak, jumlahnya dapat mencapai puluhan ribu. Oleh karena itu analisis secara manual sangat sulit dilakukan. Namun telah banyak penelitian yang menggunakan *machine learning* untuk mengatasi masalah tersebut (Fawziyyah, 2016).

Machine learning adalah salah satu disiplin dari *Artificial Intelligence* yang berfokus untuk membuat komputer mampu belajar secara otomatis tanpa harus diprogram secara eksplisit dan berulang kali. Salah satu metode dalam teknik machine learning yang digunakan dalam penelitian ini yaitu metode klasifikasi. Tujuan klasifikasi adalah untuk mengidentifikasi gen yang diekspresikan secara berbeda yang dapat digunakan untuk memprediksi keanggotaan kelas untuk sampel baru. Pada penelitian ini teknik klasifikasi dalam pembelajaran mesin yang digunakan dalam menganalisis data *Maize NAM* yaitu *K-Nearest Neighbors (KNN)* (PodFeeder, 2018).

Teknik klasifikasi yang digunakan untuk mengidentifikasi gen yang diekspresikan secara berbeda serta dapat digunakan untuk memprediksi keanggotaan kelas untuk sampel baru pada penelitian ini adalah metode klasifikasi KNN. Algoritma KNN adalah sebuah metode yang digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. (Muliadinata, 2013)

Metode KNN merupakan metode *supervised*, dimana hasil dari *query instance* yang baru akan diklasifikasikan berdasarkan mayoritas kategori yang ada pada KNN. Pada proses *training*, algoritma KNN hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi data *training sample*. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk *testing data* (klasifikasinya belum diketahui). Jarak dari vektor yang baru ini terhadap seluruh sampel vektor *training* dihitung, dan sejumlah k buah yang memiliki jarak paling dekat dengan vektor yang diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut (Muliadinata, 2013).

Klasifikasi merupakan salah satu metode pada *Machine learning*. *Machine learning* digunakan dalam sejumlah besar aplikasi dan studi bioinformatika. Menurut Thieman dan Palladino (2013) dalam Jurnal Pendekatan Bioinformatika Berbasis Penelitian Analisis Profil *Protein Carbonic Anhydrase II* yang berpotensi sebagai kandidat penyebab autisme untuk variasi pembelajaran mata kuliah bioteknologi, bioinformatika (*bioinformatics*) merupakan ilmu yang mempelajari penerapan teknik komputasional untuk mengelola dan menganalisis informasi biologis. Bidang bioinformatika mencakup penerapan metode-metode matematika, statistika, dan informatika untuk memecahkan masalah-masalah biologis, terutama dengan menggunakan sekuens DNA dan asam amino serta informasi yang berkaitan dengannya (Anisya, 2016).

Oleh karena itu, penelitian ini menggunakan salah satu metode klasifikasi pada teknik *machine learning* yang melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut yaitu metode klasifikasi KNN. Klasifikasi ini digunakan untuk mengklasifikasi nilai ekspresi gen dari berbagai galur baru tanaman jagung berdasarkan organ tanaman jagung tersebut.

1.2 Rumusan Masalah

Berdasarkan permasalahan di atas maka rumusan masalah pada penelitian tugas akhir ini adalah sebagai berikut :

1. Bagaimana aplikasi metode KNN pada data *gene expression Maize NAM*?
2. Bagaimana menentukan nilai k terbaik dari metode KNN untuk mengklasifikasikan nilai ekspresi gen pada data *gene expression Maize NAM*?
3. Bagaimana melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan KNN?
4. Bagaimana implikasi klasifikasi dengan metode KNN pada data *gene expression Maize NAM*?

1.3 Tujuan Penelitian

Berdasarkan permasalahan di atas maka tujuan penelitian ini adalah:

1. Melakukan pengklasifikasian terhadap data uji dengan menggunakan metode KNN.
2. Mendapatkan nilai k terbaik dari metode KNN untuk mengklasifikasikan nilai ekspresi gen pada data *gene expression Maize NAM*.
3. Melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan KNN.
4. Memahami implikasi metode KNN pada data *gene expression Maize NAM*.

1.4 Batasan Masalah

1. Data yang digunakan pada penelitian ini adalah data *gene expression Maize NAM*.
2. Software yang digunakan dalam penelitian ini adalah RStudio dengan menggunakan beberapa *package* bioinformatika.
3. Melakukan klasifikasi dengan menggunakan metode KNN.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Manfaat penelitian bagi penulis adalah dapat mengaplikasikan ilmu statistik yang telah didapatkan.
2. Dengan diketahui gambaran hasil klasifikasi organ pada data ekspresi gen tanaman jagung diharapkan dapat dijadikan acuan oleh penelitian lain baik dari kelebihan maupun kekurangan menggunakan KNN.