

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Sekarang ini dunia teknologi berkembang semakin pesat mempengaruhi perkembangan pada berbagai bidang lainnya. Salah satu bidang yang terpengaruh oleh perkembangan teknologi yaitu bidang pertanian. Perubahan pada bidang pertanian diawali dengan penemuan mesin dan cara baru dalam bidang tersebut. Seiring bertambahnya jumlah penduduk, semakin meningkat pula kebutuhan akan sandang, pangan dan papan sehingga peranan teknologi sangat diperlukan untuk keberhasilan produktivitas. Menurut Kementerian Pertanian (2015) untuk meningkatkan produktivitas dapat dilakukan dengan berbagai hal, seperti dengan melakukan teknik budidaya yang tepat yaitu dengan mensinergikan antara kebutuhan tanaman dalam setiap fase pertumbuhan dengan input yang harus diberikan.

Jagung (*Zea mays L.*) menjadi salah satu komoditi tanaman terpenting di dunia karena memiliki sumber karbohidrat yang tinggi. Di daerah Amerika Utara dan Selatan jagung termasuk dalam sumber karbohidrat utama dan juga sebagai salah satu alternatif sumber pangan. Menurut USDA dan NASS (2011) di Amerika Serikat pada tahun 2010 terdapat 36 juta hektar tanah ditanami tanaman jagung. Di Indonesia itu sendiri, jagung juga merupakan salah satu komoditas penting dan masih menjadi kebutuhan pangan utama masyarakat yang nilai permintaannya cukup tinggi. Pemerintah juga menjadikan jagung sebagai salah satu komoditas pangan yang akan dikembangkan (Yoninda, 2018). Selain sebagai bahan pangan, jagung juga dapat dijadikan sebagai bahan pakan ternak dan bahan baku industri. Di Indonesia dan berbagai negara lain berlomba-lomba dalam meningkatkan produksi jagung guna memenuhi permintaan industri yang tinggi. Untuk menghasilkan jagung yang bermutu tinggi, maka benih yang digunakan juga harus bermutu tinggi, baik secara genetik, fisik maupun fisiologinya dan berasal dari varietas yang unggul.

Tanaman jagung memiliki sifat-sifat vegetatif dan generatif yang perlu ditingkatkan untuk menghasilkan fenotipe yang diinginkan. Kombinasi dari keragaman genetik dan genom juga diperlukan untuk menjelaskan gen yang mendasari dan menyebarkan alel pada tanaman yang memiliki varietas yang lebih baik. Keragaman genetik itu sendiri yaitu keragaman yang ditimbulkan oleh perbedaan genetik pada setiap individu, seperti bentuk dari suatu tanaman, yaitu tinggi tanaman, umur tanaman, hasil, dan lain-lain. Sifat-sifat tersebut ditentukan oleh gen-gen tertentu pada kromosom, interaksi gen-gen antara gen dengan lingkungan (Sari, 2012).

Untuk mendapatkan jagung yang bermutu tinggi dengan varietas yang unggul yaitu dengan melakukan rekayasa genetika. Rekayasa genetika ini dilakukan dengan menyisipkan sekuens gen unggul pada suatu organisme ke organisme lain yang diinginkan melalui analisis genomik dari basis data genom organisme tersebut. Dengan berkembangnya ilmu pengetahuan dan teknologi yang berbasis pada DNA, diperlukan suatu teknologi untuk menangani itu. Salah satu teknologi itu yaitu bioinformatika.

Bioinformatika merupakan salah satu alternatif pencarian sekuens enzim baru menggunakan analisis filogeni untuk mencari spesies terdekat berdasarkan data genome yang berada pada *Gene Bank*. Bioinformatika itu sendiri adalah gabungan dari ilmu biologi, ilmu komputer, informatika, matematika, dan disiplin lain yang terkait yang menjadi disiplin tersendiri. Tujuan utama dari bioinformatika adalah dapat memberikan pandangan baru dalam mencapai perspektif global yang menunjang perkembangan bioteknologi di masa depan. Analisis dalam bioinformatika difokuskan pada tiga jenis dataset, yaitu urutan genom, struktur makromolekul dan percobaan genomik fungsional. Tetapi analisis bioinformatika juga diterapkan pada berbagai data lain, seperti pohon taksonomi, data tentang hubungan jalur metabolik, teks artikel ilmiah dan statistik. Berbagai macam teknik yang digunakan termasuk pencocokan sekuens, struktur protein 3D, konstruksi pohon filogenetik, prediksi dan klasifikasi struktur protein, prediksi struktur RNA, prediksi fungsi protein, dan ekspresi kluster data (Narita, 2014).

Salah satu penerapan ilmu bioinformatika pada bidang pertanian yaitu memanfaatkan gen pada tumbuhan untuk mendapatkan varietas yang unggul.

Salah satu teknologi yang berkembang dalam bioinformatika adalah teknologi *microarray*. *Microarray* itu sendiri merupakan sebuah perangkat berupa *chip* yang didalamnya berisi ribuan gen. Sedangkan analisis *microarray* merupakan proses analisis ekspresi gen (*gene expression*) melalui *microarray*. *Gene expression* adalah proses penerjemahan informasi dari gen menjadi produk gen yang berguna yaitu, protein (Sanchez, 2008). Teknologi *microarray* DNA adalah sebuah pendekatan untuk mempelajari ratusan hingga ribuan gen dalam suatu organisme secara bersamaan (Schena et al., 1995; DeRisi et al., 1996; Holter et al., 2000; Slonim, 2002) dalam (Dhawan, 2010).

Saat ini teknologi sekuensing DNA telah berkembang dengan munculnya *Next Generation Sequencing* (NGS). NGS merupakan salah satu teknologi sekuensing kapasitas tinggi untuk sekuensing genom total (Tasma, dkk, 2015). Metode NGS ini dapat mensekuens jutaan fragmen DNA dalam satu analisis dengan biaya yang relatif rendah dibanding teknologi sekuensing Sanger. Teknologi NGS ini sangat handal dalam mengidentifikasi variasi genom suatu spesies tanaman dalam rangka menemukan gen-gen unggul dan berbagai marka DNA dalam jumlah yang besar dalam pemuliaan tanaman (Tasman, 2014).

Mengingat perkembangan teknologi yang berkembang semakin pesat yang diikuti dengan ledakan data dalam jumlah yang sangat besar khususnya pada data biologis, maka diperlukan suatu metode yang dapat menangani permasalahan dengan jumlah data yang besar. Salah satunya yaitu dengan menggunakan data mining. Data mining merupakan teknik untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai data besar (Turban, 2005). Salah satu pendekatan data mining yaitu metode klasifikasi. Menurut Han (2006) klasifikasi adalah suatu proses untuk memperoleh model atau fungsi yang menjelaskan dan membedakan konsep atau kelas data dengan tujuan untuk memprediksi suatu kelas dari objek yang tidak diketahui labelnya. Metode klasifikasi yang digunakan peneliti dalam penelitian ini yaitu klasifikasi *K-Nearest Neighbors* (KNN).

KNN merupakan sebuah metode yang digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN termasuk dalam klasifikasi yang sederhana namun efektif digunakan pada klasifikasi pola. Tujuan dari metode ini adalah mengklasifikasikan objek baru berdasarkan atribut dan *training sample*. Karena kesederhanaan dan popularitasnya, metode KNN ini sudah digunakan secara luas dalam pengenalan pola, pembelajaran mesin, kategorisasi teks, data mining, pengenalan objek, dll. (Alkasassbeh, 2015).

Penelitian ini bertujuan untuk mengklasifikasikan data gen tanaman jagung berdasarkan nilai *expression set* dari tanaman jagung yang telah diperoleh dengan menggunakan KNN.

## 1.2 Rumusan Masalah

Berdasarkan dari latar belakang di atas, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana aplikasi metode KNN pada data ekspresi gen *Maize Growth Stages*?
2. Bagaimana menentukan nilai  $k$  terbaik dari metode KNN untuk mengklasifikasikan tahap pertumbuhan pada organ jagung (*Maize Growth Stages*) menggunakan metode KNN?
3. Bagaimana melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan metode KNN?
4. Bagaimana implikasi metode KNN pada data *Maize Growth Stages*?

## 1.3 Batasan Masalah

Agar penelitian tidak menyimpang dari pembahasan maka perlu diberikan suatu pembatas masalah yaitu sebagai berikut.

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

#### **1.4 Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah sebagai berikut.

1. Melakukan penerapan metode KNN pada data ekspresi gen *Maize Growth Stages*.
2. Mendapatkan nilai  $k$  terbaik dari metode KNN untuk mengklasifikasikan tahap pertumbuhan pada organ jagung (*Maize Growth Stages*) menggunakan metode KNN.
3. Melakukan evaluasi terhadap hasil klasifikasi dengan metode KNN.
4. Memahami implikasi metode KNN pada data *Maize Growth Stages*.

#### **1.5 Manfaat Penelitian**

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Manfaat bagi penulis adalah pengaplikasian ilmu statistik yang telah didapatkan selama kuliah.
2. Hasil penelitian ini mampu memberikan gambaran kepada berbagai kalangan mengenai hasil klasifikasi organ jagung pada data ekspresi gen dan dapat memprediksi kelompok dari data yang baru.