

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada jurnal penelitian terdapat banyak jenis kasus yang bisa diangkat, Salah satunya adalah peng-klasifikasian. Klasifikasi merupakan salah satu kasus yang banyak diangkat oleh peneliti, hal ini ditunjukkan dari banyaknya jurnal yang menggunakan kasus ini. Untuk melakukan klasifikasi kita ditawarkan dengan banyaknya metode, yang dimana metode-metode tersebut memiliki ciri dan hasil yang berbeda-beda.

Untuk melakukan klasifikasi terhadap data maupun gambar ada banyak metode yang dapat diterapkan, diantaranya yaitu Decission Tree, Naïve Bayes, K-Nearest Neighbor, dan Suppot Vector Machine. Decission tree adalah metode yang dapat difungsikan untuk melakukan klasifikasi terhadap banyak data. Metode ini dapat mengelolah data yang besar dan membuat pohon keputusan[1]. K-Nearest Neighbor (K-NN) adalah metode yang dapat digunakan untuk melakukan klasifikasi kepada objek berdasarkan ciri-ciri data latih yang paling dekat dengan objeknya. Metode ini menghitung jarak sehingga bisa diolah menjadi hitungan matematis dengan teknik perhitungan jarak Euclidean Distance[11]. Naïve Bayes adalah teknik pengklasifikasian data probabilistic sederhana yang cara kerjanya menghitung sekumpulan probabilitas, frekuensi dan kombinasi dari data. Dengan menggunakan teorema bayes dan menganggap semua attribute adalah independen atau tidak ketergantungan antar data[21]. Support Vector Machine (SVM) adalah metode diskriminatif yang paling baik dilakukan dalam klasifikasi data. Metode ini bekerja pada structural risk minimation yang adalah prinsip dari pembelajaran mesin[32].

Banyaknya jurnal yang sudah diterbitkan dengan menggunakan ke 4 metode tersebut dalam melakukan klasifikasi. membuat penulis merasa penasaran terhadap metode yang paling baik untuk melakukan klasifikasi terhadap data. Sehingga penulis akhirnya tertarik untuk membawa judul berupa “PERBANDINGAN EFEKTIFITAS ALGORITMA DECISSION TREE, NAÏVE BAYES, K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE DALAM MELAKUKAN KLASIFIKASI”

1.2 Rumusan Masalah

Berdasar dari latar belakang yang ada diatas, maka dapat dirumuskan permasalahan saat ini adalah apakah metode yang paling cocok untuk melakukan peng-klasifikasian data.

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk menemukan metode yang menghasilkan efektifitas terbaik dalam melakukan klasifikasi terhadap data.

1.4 Manfaat Penelitian

Adapun beberapa manfaat dari penelitian ini adalah sebagai berikut :

1. Membantu masyarakat dalam menemukan metode yang tepat pada penelitian yang akan dilakukannya.
2. Memudahkan masyarakat dalam menilai hasil akurasi rata-rata terhadap metode yang akan diteliti.
3. Membantu masyarakat dalam menilai metode apa yang paling tepat dalam melakukan klasifikasi.

1.5 Batasan Masalah

Adapun batasan-batasan masalah dalam penelitian ini adalah sebagai berikut :

1. File data didapatkan dengan mendownload manual dari website google scholar.
2. Metode yang akan dibandingkan adalah 4 metode dalam melakukan klasifikasi yaitu, Decission Tree, Naïve bayes, K-Nearest Neighbor dan Support Vector Machine.
3. Setiap metode terdiri dari 10 jurnal sehingga total ada 40 data jurnal.

1.6 Keterbaruan

Klasifikasi adalah sebuah kegiatan untuk melakukan analisa data yang menciptakan model-model untuk menggambarkan kelas-kelas yang didapati pada data tersebut. Dalam melakukan klasifikasi ada banyak metode yang dapat digunakan, tapi pada penelitian ini hanya akan membandingkan 4 buah metode saja.

1. Menurut Imam Sutoyo, (2018), pada jurnal dengan judul Implementasi Algoritma Decision Tree Untuk Melakukan Klasifikasi Data Peserta Didik, menyatakan bahwa klasifikasi terhadap peserta didik yang memiliki tujuan untuk melakukan pengelompokkan peserta agar kegiatan belajar mengajar dapat disesuaikan. Metode sederhana untuk melakukan klasifikasi adalah dengan mengurutkan peserta didik dengan menggunakan 1 buah attribute[3].
2. Menurut Cinantya Paramita, dkk., (2019), pada jurnal yang berjudul Klasifikasi dari jeruk Nipis Terhadap Tingkat Kematangan Buah yang didasarkan dari Fitur Warna Menggunakan K-Nearest Neighbor, menyatakan bahwa klasifikasi terhadap jeruk nipis dahulu dilakukan secara manual dengan mata manusia merupakan hal yang sulit untuk dilakukan. Hal ini disebabkan karena ketidak konsistenan serta mata manusia yang bersifat subyektif sehingga menghasilkan efektifitas yang kecil[12].
3. Menurut Riri Nada Devita, dkk., (2018), pada jurnal yang berjudul Perbandingan Dari Kinerja Metode Naïve Bayes Dan K-Nearest Neighbor Untuk klasifikasi Artikel Yang Berbahasa Indonesia, menyatakan bahwa ada banyak metode klasifikasi yang bisa digunakan. Namun pada penelitian ini digunakan Naïve Bayes dan K-Nearest Neighbor. Pemilihan metode Naïve Bayes dipilih karena dapat menghasilkan nilai akurasi yang maksimal walau hanya memiliki data latih hanya sedikit. Sedangkan penggunaan K-Nearest Neighbor digunakan karena metode ini memiliki nilai toleransi terhadap noise yang tinggi[23].
4. Menurut Ana Mariyam Puspitasari, dkk., (2018), pada jurnal yang berjudul Klasifikasi Dari Penyakit Gigi Dan Mulut Dengan Menggunakan Metode Support Vector Machine, menyatakan bahwa pada penelitian ini digunakan metode Support Vector Machine dikarenakan metode ini mampu mengatasi masalah klasifikasi dan regresi secara linear maupun non-linear[31].