

PERBANDINGAN K-NEAREST NEIGHBOR DAN NAIVE BAYES UNTUK KLASIFIKASI TANAH LAYAK TANAM POHON JATI

Didik Srianto¹, Edy Mulyanto²

^{1,2}Teknik Informatika, Universitas Dian Nuswantoro

Jl Imam Bonjol No.207 Semarang 50131,(024) 3517261

Email : 111201106069@mhs.dinus.ac.id¹, lumyde007@gmail.com²

Abstrak

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, data mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field dari sebuah relasional database yang besar. Pada perum perhutani KPH SEMARANG saat ini masih menggunakan cara manual untuk menentukan jenis tanaman (jati / non jati). K-Nearest Neighbour atau k-NN merupakan algoritma data mining yang dapat digunakan untuk proses klasifikasi dan regresi. Naive bayes Classifier merupakan suatu teknik yang dapat digunakan untuk teknik klasifikasi. Pada penelitian ini k-NN dan Naive Bayes akan digunakan untuk mengklasifikasi data pohon jati dari perum perhutani KPH SEMARANG. Yang mana hasil klasifikasi dari k-NN dan Naive Bayes akan dibandingkan hasilnya. Pengujian dilakukan menggunakan software RapidMiner. Setelah dilakukan pengujian k-NN dianggap lebih baik dari Naive Bayes dengan akurasi 96.66% dan 82.63.

Kata kunci -k-NN,Klasifikasi,Naive Bayes,Penanaman Pohon Jati

Abstract

Data mining is the process of analyzing data from different perspectives and conclude it becomes important information that can be used to increase profits, reduce costs, or both. Technically, data mining can be referred to as a process of finding correlations or patterns of hundreds or thousands of field of a large relational database. Perum Perhutani KPH on SEMARANG today still use manual to determine the type of plant (identity / non-identity). K-Nearest Neighbour or k-NN is a data mining algorithm that can be used for classification and regression. Naive Bayes Classifier is a technique that can be used for classification techniques. In this study, k-NN and Naive Bayes will be used to classify the data of Perum Perhutani teak KPH SEMARANG. Which classification results of k-NN and Naive Bayes will compare the results. Testing is done using software RapidMiner. After testing the k-NN is considered better than Naive Bayes with an accuracy of 96.66% and 82.63.

Keywords -k-NN, Klasifikasi, Naive Bayes, Teak Tree

1. PENDAHULUAN

Perum Perhutani merupakan Badan Usaha Milik Negara di Indonesia yang memiliki tugas dan wewenang untuk menyelenggarakan perencanaan, pengurusan, pengusahaan dan perlindungan hutan di wilayah kerjanya. Saat ini dasar hukum yang mengatur Perum Perhutani adalah

Peraturan Pemerintah Nomor 72 tahun 2010. Salah satu misi perum perhutani adalah mengelola sumberdaya hutan secara lestari dengan cara penanaman pohon pohon tertentu sesuai dengan kondisi tanah pada daerah yang hendak ditanam. Divisi Regional Jawa Tengah dari Perum Perhutani terdiri dari 20 Kesatuan Pemangkuan Hutan (KPH), salah satunya yaitu KPH Semarang.

Pada KPH Semarang jenis pohon yang paling sering dikelola adalah jenis pohon jati. Pemilihan pohon jati ini dikarenakan pori-pori kayunya lebih padat, teksturnya lebih berminyak, warna kayunya lebih hidup, sangat awet. Selama ini penentuan jenis pohon jati untuk ditanam pada kondisi tanah tertentu pada Perum Perhutani KPH Semarang masih menggunakan cara yang manual. Perkembangan ilmu teknologi saat ini telah bisa dimanfaatkan untuk membantu dalam mengklasifikasi penanaman pohon jati. Salah satunya dengan menggunakan bidang ilmu data mining.

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar [1]. Dalam data mining terdapat beberapa metode, diantaranya yaitu clustering, regresi dan klasifikasi. Terdapat beberapa algoritma yang dapat digunakan untuk melakukan klasifikasi, dua diantaranya yaitu k-Nearest Neighbour (k-NN) dan naïve bayes.

k-Nearest Neighbour atau k-NN merupakan algoritma data mining yang dapat digunakan untuk proses klasifikasi dan regresi. k-NN pernah diterapkan untuk merancang sebuah sistem perangkat lunak yang dapat digunakan untuk mengklasifikasikan motif batik [2]. Proses penerapannya terdiri dari input Ekstraksi ciri Local Binary Pattern (LBP), Citra, preprocessing, pembangunan model klasifikasi k-NN, serta proses klasifikasi

k-NN. Dalam penelitian tersebut hasil akurasi yang diperoleh menggunakan algoritma k-NN menunjukkan persentase sebesar 99,14%.

Naïve bayes Classifier merupakan suatu teknik yang dapat digunakan untuk teknik klasifikasi. Algoritma Naïve bayes pernah digunakan untuk mengklasifikasi teks untuk mengelompokkan teks berita dan abstrak akademis [3]. Dalam penelitian tersebut disebutkan bahwa naïve bayes memiliki kelebihan dalam kesederhanaan komputasi. Hasil yang diperoleh dari klasifikasi menggunakan naïve bayes ini menunjukkan akurasi sebesar 91%.

Berdasarkan pertimbangan diatas, pada penelitian ini akan dilakukan analisis untuk klasifikasi penanaman pohon jati di Kesatuan Pemangkuan Hutan Semarang menggunakan algoritma k-NN dan naïve bayes.

2. METODE PENELITIAN

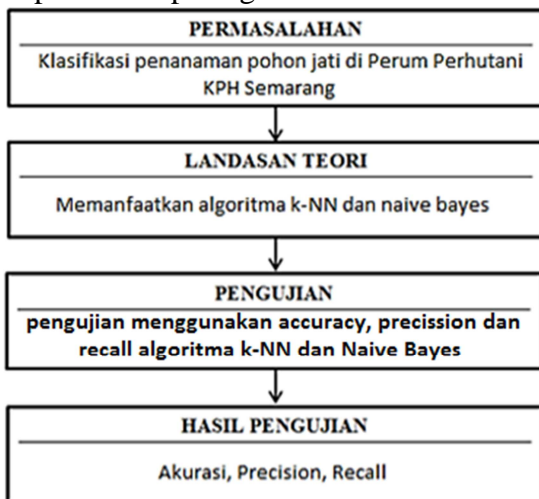
Pada penelitian ini akan dilakukan tahapan-tahapan yang meliputi [4][5]:

1. Penggunaan dataset yang mana dataset tersebut berasal dari Perum Perhutani Kesatuan Pemangkuan Hutan Semarang, berupa dataset "Penanaman Pohon Jati".
2. Penentuan algoritma yang digunakan. Adapun algoritma yang akan digunakan dalam penelitian ini yaitu algoritma K-NN dan Naïve Bayes.
3. Implementasi, implementasi pada penelitian ini akan menggunakan software RapidMiner 5.3.
4. Pengujian, untuk pengujian pada algoritma K-NN akan dilakukan 5 kali dengan memberikan nilai K yang berbeda pada setiap pengujian yaitu

1, 3, 5, 7 dan 9. Sedangkan untuk Naïve Bayes akan menggunakan teknik *split validation* dan pengujian akan dilakukan sebanyak 5 kali dengan setiap pengujian dibagi menjadi data *training* dan data *testing* yaitu 95% & 5%, 85% & 15%, 75% & 25%, 65% & 35%, 55% & 45%.

5. Analisis pengujian, yang akan dianalisis pada penelitian ini adalah *accuracy*, *precision*, *recall* dari masing-masing algoritma.
6. Hasil pengujian, dalam penelitian ini akan dapat diketahui algoritma mana yang memiliki akurasi paling tinggi untuk dataset penanaman pohon jati.

Adapun tahapan pada penelitian ini dapat dilihat pada gambar 1 :

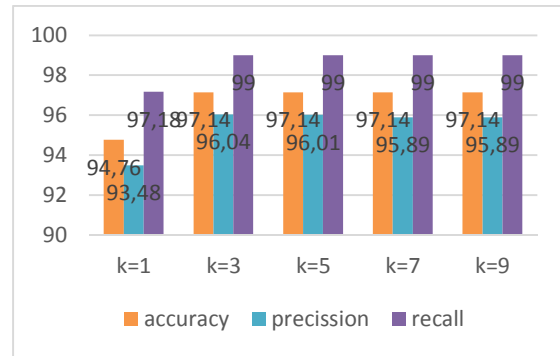


Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengujian K-NN

Berdasarkan hasil pengujian menggunakan *RapidMiner 5.3*, maka diperoleh *accuracy*, *precision*, dan *recall* k-NN sebagai berikut[6][7] :



Gambar 2 . Diagram Accuracy, Precision, Recall k-NN

Pada gambar 2 menunjukkan hasil *accuracy* untuk k=1 sebesar 94.76%, k=3 sebesar 97.14%, k=5 sebesar 97.14%, k=7 sebesar 97.14%, dan k=9 sebesar 97.14%. Hasil untuk *precision* diperoleh k=1 sebesar 93.48%, k=3 sebesar 96.04%, k=5 sebesar 96.01%, k=7 sebesar 95.89%, dan k=9 sebesar 95.89%. Hasil untuk *recall* diperoleh k=1 sebesar 97.18%, k=3 sebesar 99%, k=5 sebesar 99%, k=7 sebesar 99%, dan k=9 sebesar 99%. Adapun secara keseluruhan hasil pengujian k-NN dapat dilihat pada tabel 1.

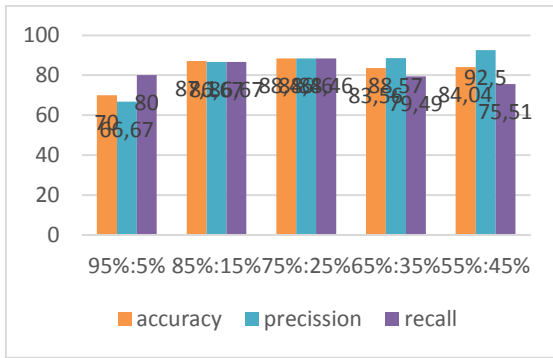
Tabel 1: Hasil Pengujian k-NN

Nilai k	Accuracy (%)	Precision(%)	Recall(%)
k=1	94.76	93.48	97.18
k=3	97.14	96.04	99
k=5	97.14	96.01	99
k=7	97.14	95.89	99
k=9	97.14	95.89	99

Apabila hasil yang diperoleh diambil rata-rata maka diperoleh *accuracy* sebesar 96.66%, *precision* sebesar 95.46%, dan *recall* sebesar 98.63%.

3.2 Hasil Pengujian Naïve Bayes

Berdasarkan hasil pengujian menggunakan *RapidMiner 5.3*, maka diperoleh *accuracy*, *precision*, dan *recall* Naïve Bayes sebagai berikut :



Gambar 3. Diagram Accuracy, Precision, Recall Naïve Bayes

Pada gambar 3 menunjukkan hasil *accuracy* untuk *training* = 95% dan *testing* = 5% sebesar 70%, *training* = 85% dan *testing* = 15% sebesar 87.1%, *training* = 75% dan *testing* = 25% sebesar 88.56%, *training* = 65% dan *testing* = 35% sebesar 83.56%, dan *training* = 55% dan *testing* = 45% sebesar 84.04%. Hasil *precision* untuk *training* = 95% dan *testing* = 5% sebesar 66.67%, *training* = 85% dan *testing* = 15% sebesar 86.67%, *training* = 75% dan *testing* = 25% sebesar 88.56%, *training* = 65% dan *testing* = 35% sebesar 88.57%, dan *training* = 55% dan *testing* = 45% sebesar 92.5%. Hasil *recall* untuk *training* = 95% dan *testing* = 5% sebesar 80%, *training* = 85% dan *testing* = 15% sebesar 86.67%, *training* = 75% dan *testing* = 25% sebesar 88.56%, *training* = 65% dan *testing* = 35% sebesar 79.49%, dan *training* = 55% dan *testing* = 45% sebesar 75.51%. Adapun secara keseluruhan hasil pengujian Naïve Bayes dapat dilihat pada tabel 2.

Tabel 2: Hasil Pengujian Naïve Bayes

	k-NN	Naïve Bayes	Unggul
Accuracy(%)	96.66	82.63	k-NN
Precision(%)	95.46	84.57	k-NN
Recall(%)	98.63	82.02	k-NN

Apabila hasil yang diperoleh diambil rata-rata maka diperoleh *accuracy* sebesar 82.63%, *precision* sebesar 84.57%, *recall* sebesar 82.2%.

3.3 Analisis Pengujian

Dari pengujian pada algoritma k-NN dan Naïve Bayes menggunakan dataset penanaman pohon jati yang sudah dilakukan maka diperoleh perbandingan antara algoritma k-NN dan Naïve Bayes (NB) yang dapat dilihat pada tabel 3.

Tabel 3: Analisis Pengujian k-NN dan Naïve Bayes

Training : Testing	Accuracy (%)	Precision(%)	Recall(%)
95%:5%	70	66.67	80
85%:15%	87.1	86.67	86.67
75%:25%	88.46	88.46	88.46
65%:35%	83.56	88.57	79.49
55%:45%	84.04	92.5	75.51

Dari tabel 3 dapat diketahui bahwa algoritma k-NN lebih unggul disemua bagian, baik *accuracy*, *precision* dan *recall*.

4. KESIMPULAN

Hasil pengujian menunjukkan bahwa pada penelitian ini diperoleh *accuracy* k-NN lebih tinggi dari pada *accuracy* Naive Bayes dengan perbandingan 96.66 (k-NN) dan 82.63% (Naïve Bayes). Sedangkan pada bagian *precision* menunjukkan hasil k-NN yang lebih tinggi dari pada Naïve Bayes dengan perbandingan 95.45% (k-NN) dan 84.57% (Naïve Bayes). Pada bagian *recall*, k-NN kembali memperoleh hasil yang lebih tinggi dari pada Naïve Bayes dengan perbandingan 98.63 (k-NN) dan 82.02% (Naïve Bayes). Sehingga dapat dianggap bahwa klasifikasi menggunakan algoritma k-NN lebih baik dibandingkan algoritma Naïve

bayes untuk mengklasifikasi data penanaman pohon jati dari KPH Semarang.

5. SARAN

Dalam penerapan algoritma k-NN dan Naive Bayes untuk klasifikasi penanaman pohon jati dari KPH Semarang, terdapat beberapa hal yang perlu diperhatikan agar menjadi lebih baik kedepannya, diantaranya sebagai yaitu :

1. Menambahkan fitur seleksi pada proses klasifikasi untuk memperoleh hasil yang lebih optimal.
2. Dataset dan atribut yang digunakan dapat menggunakan jumlah data dan atribut yang lebih banyak dan kompleks.

DAFTAR PUSTAKA

- [1] Y. Religia, "Analisis Algoritma k-NN dan Naive Bayes Untuk Klasifikasi Pembelian Mobil," UDINUS, Semarang, 2015.
- [2] H. Wahyu, I. I. Tritasmoro and R. D. Atmaja, "Perancangan Sistem Perangkat Lunak Untuk Mengklasifikasikan Motif Batik Menggunakan Metode K-Nearest Neighbour (K-NN)," Telkom University, 2013.
- [3] A. Hamzah, "Klasifikasi Teks Dengan Naive Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita Dan Abstract Akademis," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, pp. 269-277, 3 Nopember 2012.
- [4] C. Parida Purnama, "Deteksi Penyakit Diabetes Type II Dengan Naive Bayes Berbasis Particle Swarm Optimization," *jurnal teknologi informasi*, vol. 9, no. 2, 2013.
- [5] S. mujahidin, "Klasifikasi Warna Kulit Berdasarkan Ruang," *seminar nasional aplikasi teknologi informasi (SNATI)*, 2015.
- [6] S. Riyan Eko Putri, "Perbandingan Metode Klasifikasi Naive Bayes dan k-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012," *jurnal gaussian*, vol. 3, no. 4, pp. 831-838, 2014.
- [7] D. Rahmawati. Indriati. Santosa Dedy, "Perbandingan Kinerja Metode Naive Bayes, K-Nearest Neighbors, dan Metode Gabungan K-Mans dan LVQ Dalam Pengkategorian Buku Komputer Berbahasa Indonesia Berdasarkan Judul dan Sinopsis".