ARTIKEL

ANALISA PERBANDINGAN METODE KLASIFIKASI NAIVE BAYES DENGAN LAPLACIAN DAN K-NEAREST NEIGHTBOR (K-NN) PADA SELEKSI PENERIMA BEASISWA

Oleh:

DEWI RAHMAWATI

NPM : 13.1.03.02.0193

Dibimbing oleh :

Ahmad Bagus Setiawan, S.T., M.M., M.Kom
Fajar Rohman Hariri, M.Kom

FAKULTAS TEKNIK (FT)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
2017

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2017

Yang bertanda tangan di bawah ini:

Nama Lengkap : DEWI RAHMAWATI
NPM : 13.1.03.02.0193
Telepun/HP : 085 649 400 81
Alamat Surel (Email) : rahmadewi951@gmail.com
Judul Artikel : Analisa Perbandingan Metode Klasifikasi Naive Bayes dengan Laplacian dan K-Nearest Neightbor (K-NN) Pada Seleksi Penerima Beasiswa
Fakultas – Program Studi : Teknik – Teknik Informatika
Nama Perguruan Tinggi : Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi : Jln. K.H Ahmad Dahlan No 76 Kediri

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme.
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggung jawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 08 Agustus 2017
Pembimbing I	Pembimbing II	Penulis,
Ahmad Bagus Setiawan, S.T., M.M.,M.Kom. NIDN. 0703018704	Fajar Rohman Hariri, M.Kom NIDN. 0715058902	Dewi Rahwawati NPM. 13.1.03.02.0193

ANALISA PERBANDINGAN METODE KLASIFIKASI NAIVE BAYES DENGAN LAPLACIAN DAN K-NEAREST NEIGHTBOR (K-NN) PADA SELEKSI PENERIMA BEASISWA

DEWI RAHMAWATI
13.1.03.02.0193
Teknik – Teknik Informatika
Rahmadewi951@gmail.com
Ahmad Bagus Setiawan, S.T., M.M.,M.Kom.dan Fajar Rohman Hariri, M.kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

Abstrak

DEWI RAHMAWATI: Analisa Perbandingan Metode Klasifikasi *Naive Bayes* Dengan *Laplacian* Dan *K-Nearest Neightbor* (K-NN) Pada Seleksi Penerima Beasiswa, Skripsi, Teknik Informatika, FT UN PGRI Kediri, 2017.

Penelitian ini dilatar belakangi permasalahan yang sering muncul dalam penyaluran beasiswa, yaitu kurang tepatnya penyaluran beasiswa karena tujuan dari adanya beasiswa adalah memotivasi siswa untuk selalu meningkatkan prestasi akademik maupun non akademik. Dari permasalahan tersebut banyak dibuat sistem penerima beasiswa yang bertujuan mempermudah pihak sekolah untuk menyeleksi penerima beasiswa dengan berbagai metode. Banyaknya metode yang digunakan dengan hasil yang berbeda-beda membuat beberapa peneliti melakukan analisa metode mana yang lebih akurat digunakan untuk sistem penerima beasiswa.

Permasalahan penelitian ini adalah (1) Bagaimana menganalisa hasil perbandingan metode klasifikasi *naive bayes* dengan *laplacian* dan *K-Nearest Neightbor (K-NN)* terhadap sistem seleksi penerima beasiswa?

Penelitian ini menggunakan dua metode klasifikasi yaitu *naive bayes* dengan *laplacian* dan *K-Nearest Neightbor (K-NN). Naive bayes* dengan *laplacian* yaitu klasifikasi dengan memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya dengan menambahkan nilai 1 pda setiap probabilitas untuk mencegah nilai nol. Sedangkan, *K-Nearest Neightbor (K-NN)* adalah metode klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Kesimpulan dari penelitian ini adalah dari hasil perhitungan dan pengujian sistem perbandingan metode pada seleksi penerima beasiswa didapat bahwa perhitungan *naive bayes* dengan *laplacian* menghasilkan 5 siswa yang mendapatkan beasiswa dari 13 siswa. Sedangkan, metode k-nn (*k-nearest neigtbor)* menghasilkan 3 siswa yang mendapatkan beasiswa dari 13 siswa. Itu artinya metode *naive bayes* dengan *laplacian* lebih tepat digunakan untuk sistem penerima beasiswa.

Kata Kunci : Perbandingan, *naive bayes* dengan *laplacian, K-Nearest Neightbor (K-NN),* penerima beasiswa, Hasil Analisa.

I. LATAR BELAKANG

Dalam rangka meningkatkan akses dan minat belajar siswa serta mengangkat mutu sekolah, SMPN 3 Plosoklaten mengalokasikan dana beasiswa dalam bentuk bantuan untuk siswa yang kurang mampu dan berprestasi. Namun masih ada permasalahan yang sering muncul, yaitu kurang tepatnya penyaluran beasiswa. Dari permasalahan tersebut banyak dibuat sistem penerima beasiswa yang bertujuan mempermudah pihak sekolah untuk menyeleksi penerima beasiswa dengan berbagai metode. Banyaknya metode yang digunakan membuat beberapa peneliti melakukan analisa metode mana yang lebih akurat digunakan untuk sistem penerima beasiswa.

Berdasarkan permasalahan tersebut dapat diambil solusi dengan cara membuat sistem dengan membandingkan 2 metode untuk seleksi calon penerima beasiswa berdasarkan prestasi yaitu dengan menggunakan metode *naive bayes clasification* dengan *laplasian* dan *K-Nearest Neightbor (KNN).*

Metode *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian.

Sedangkan, Metode *K-Nearest Neightbor (K-NN)* adalah sebuah metode 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.

Dengan adanya sistem perbandingan metode ini diharapkan dapat membantu menemukan informasi tentang metode mana yang lebih tepat

dan akurat untuk digunakan dalam studi kasus penerima beasiswa.

II. METODE

1. *Naive Bayes Classification*

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma mengunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output.

Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan. (Saleh, 2015)

Persamaan Metode *Naive Bayes* adalah :

$$P(H|X) = \frac{\mathrm{P(X|H).P(H)}}{\mathrm{P(X)}} \quad .....(1)$$

2. *Laplace Estimator*

Untuk menyiasati supaya hasil probabilitas pada perhitungan Naïve Bayes tidak bernilai nol dikarenakan tidak adanya data untuk suatu kategori tertentu dalam kelasnya dapat digunakan teknik *estimasi* yang biasa disebut *Laplace estimator* atau *Laplacian correction.* Dalam teknik ini digunakan penambahan nilai 1 pada data untuk masing-masing kategori ketika ada kategori yang memiliki nilai 0 sehingga untuk sebanyak k kategori dimana j=1,2,...,k dan N= $N \sum_{j-1}^{k}$ nj jika

masing-masing kategori dalam kelasnya bernilai n maka : (Riyan dkk, 2014)

$$P(X=i)\ \frac{ni+1}{n+\text{banyak kategori}} \text{..... (3)}$$

3. *K-Nearest Neighbor (K-NN)*

KNN (*K-Nearest Neighbor*) merupakan metode yang cukup popular dan sederhana. KNN termasuk metode klasifikasi data mining yang didasarkan pada pembelajaran dengan analogi. Sampel data pelatihan memiliki *n* atribut dimensi numerik. Setiap sampel merupakan titik dalam ruang *n*- dimensi. Semua sampel pelatihan disimpan di ruang *n*-dimensi. Ketika pengujian data, akan mencari nilai *k* terdekat dengan data uji. Kedekatan didefinisikan dalam hal jarak *Euclidean* antara dua titik. $X = (X_1, X_2,..X_n)$ dan $Y = (Y_1,Y_2,..Y_n)$ ....persamaan(1)

merupakan persamaan yang digunakan pada metode KNN : (Mega dkk, 2015)

$$d(X,Y)= \sqrt{\sum_{i=1}^{n}(xi - yi)2} \text{.....(2)}$$

Tahapan analisis dengan metode *K-Nearest neighbor* (Ririn dkk, 2014)

1. Menentukan kelas
   Kelas variabel dependen berupa kategorikalyangmerepresentasikan ‘label’ yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa.
2. *Predictor*
   Variabelindependendirepresentasikan oleh karakteristik (atribut) data. Contohnya : merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
3. *Training dataset*
   Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan prediktor.
4. *Testing dataset*
   Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

5. Pengukuran akurasi

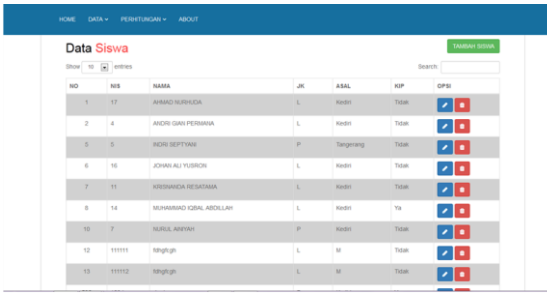
Untuk mengukur akurasi algoritma klasifikasi, metode yang dapat digunakan yaitu *cross validation*, *confusion matrix* dan kurva ROC (*ReceiverOperatingCharacteristic*) Untuk mengembangkan aplikasi (*development*) berdasarkan model yang dibuat, digunakan *Rapid Miner*. Sedangkan metode validasi yang sering digunakan adalah *confusion matrix.*

III. HASIL DAN KESIMPULAN

1. Hasil

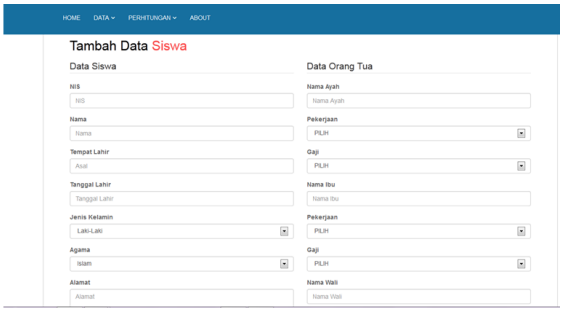
a. Tampilan Data Siswa

Pada halaman ini digunakan untuk melihat data siswa baru.didalamnya juga terdapat menu edit, tambah, hapus.



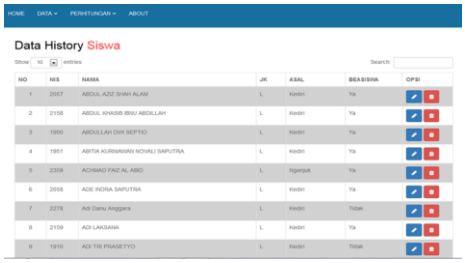
b. Menu Input Data Siswa

Pada halaman ini terdapat menu untuk menambahkan data siswa yang baru. Mulai data diri siswa maupun data orang tua atau wali siswa.



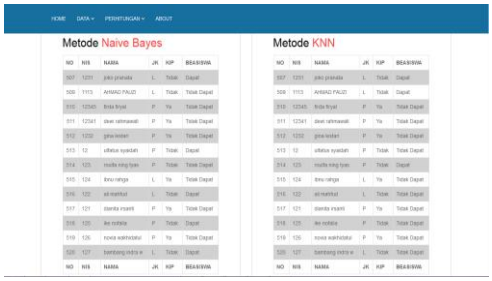
c. Tampilan Data History

Terdapat data siswa yang pernah mendapatkan maupun tidak mendapatkan beasiswa digunakan sebagai acuan untuk menghitung dua metode.



d. Hasil

Terdapat tampilan untuk menampilkan hasil dari perhitungan dua metode klasifikasi yaitu *naive bayes* dengan *laplacian* dan K-NN



e. Hasil Perbandingan

Pada menu ini terdapat dari hasil perbandingan dua metode.

518	125	ike nofalia	P	Tidak	Dapat
519	126	novia wakhidatul	P	Ya	Tidak Dapat
520	127	bambang indra w	L	Tidak	Dapat
NO	NIS	NAMA	JK	KIP	BEASISWA

Siswa yang dapat 5 dari 13 siswa

518	125	ike nofalia	P	Tidak	Tidak Dapat
519	126	novia wakhidatul	P	Ya	Tidak Dapat
520	127	bambang indra w	L	Tidak	Tidak Dapat
NO	NIS	NAMA	JK	KIP	BEASISWA

Siswa yang dapat 3 dari 13 siswa

2. Kesimpulan

Berdasarkan hasil penelitian penulis dapat memberikan kesimpulan bahwa pemilihan metode terbaik adalah metode yang meghasilkan banyak siswa yang mendapatkan beasiswa yang mempunyai tingkat nilai yang lebih besar. Dari data yang terangkum, metode yang lebih baik untuk sistem penerima beasiswa yaitu metode *Naive Bayes* dengan *lapalcian* dengan hasil yang diperoleh adalah bahwa perhitungan *naive bayes* dengan *laplacian* menghasilkan 5 siswa yang mendapatkan beasiswa dari 13 siswa.

3. Saran

Saran yang dapat peneliti sampaikan antara lain :

a) Dapat menggunakan studi kasus lain dalam perbandingan dua metode.
b) Dapat menggunakan aplikasi lain dalam pembuatan aplikasi perbandingan dua metode.
c) Dapat menggunakan metode lain untuk perbandingan metode pada sistem penerima beasiswa

IV. DAFTAR PUSTAKA

Andriyana, V. 2015. Perbandingan 3 Metode Dalam *Data Mining* Untuk Prediksi Penerima Beasiswa Berdasarkan Prestasi Di Sma Negeri 6 Surakarta, Universitas Muhammadiyah Surakarta**.**

Haryana, KM. S. 2008. Pengembangan Perangkat Lunak dengan Menggunakan PHP , Jurnal Computech & Bisnis**,** Vol. 2, No. 1, Juni 2008, 14-21.

Jayanti, R. D., Noeryanti, 2014. Aplikasi Metode K-Nearest Neighbor Dan Analisis Diskriminan Untuk Analisis Resiko Kredit Pada Koperasi Simpan Pinjam Di Kopinkra Sumber Rejeki, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November2014.

Putri, R.E., Suparti, Rahmawati, R., 2014. Perbandingan Metode Klasifikasi *Naïve Bayes* Dan *K-Nearest Neighbor* Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012 , JURNAL GAUSSIAN, Volume 3, Nomor 4, Tahun 2014, Halaman 831 – 838.

Saleh, A. 2015. Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, Citec Journal, Vol. 2, No. 3, Mei 2015 – Juli 2015.

Sari, M. K., Ernawati, Pranowo, 2015. Kombinasi Metode *K-Nearest Neighbor* Dan *Naive Bayes* Untuk Klasifikasi Data, Seminar Nasional Teknologi Informasi dan Multimedia 2015 STMIK AMIKOM Yogyakarta, 6-8 Februari 2015.

Shella, P. 2015 Sistem Pendukung Keputusan Dengan Menggunakan Decision Tree Dalam Pemberian Beasiswa disekolah Menengah Pertama (Studi Kasus di SMP N 2 Ngrembang). Universitas Negeri Semarang : Semarang.

Supriyanti, W., Kusrini, Amborowati, A. 2016. Perbandingan Kinerja Algoritma C4.5 Dan *Naive Bayes* Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa, Jurnal INFORMA Politeknik Indonusa Surakarta ISSN : 2442-7942 Vol. 1 Nomor 3 Tahun 2016.

Trimarsiah, Y. 2014 Rancang Bangun Sistem Informasi Berbasis Web Pada Privat Awal Belajar Cerdas Menggunakan PHP dan MYSQL, Jurnal Media Informatika dan Komputer, Vol. 4 No.1 Desember 2014.