JURNAL INFORMATIKA, Vol.7 No.1 April 2020, Halaman 55~61
ISSN: 2355-6579 | E-ISSN**:** 2528-2247

Komparasi Algortima C4.5, *Naïve Bayes,* dan *k-Nearest Neighbor* Sebagai Sistem Pendukung Keputusan Menaikkan Jumlah Peserta Didik

Harianto[1], Didi Rosiyadi[2]

[1,2] Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
e-mail : [1] rian2x@yahoo.co.id, [2] didi.rosiyadi@gmail.com

Abstrak

Persaingan dalam menarik minat orang tua siswa untuk menyekolahkan anak mereka ke sekolah swasta memang sangat tinggi. Sekolah swasta harus bekerja keras dalam hal mendapatkan siswa yang akan melanjutkan pendidikannya ke tingkat selanjutnya. Sekolah hendaknya harus memiliki nilai tambah yang lebih di mata orang tua siswa. Dari ketiga algoritma yang digunakan, terdapat nilai accuracy tertinggi sebesar 86,50% dihasilkan dari algoritma Naïve Bayes. Dalam artian, tingkat prediksi untuk sekolah SMP Cenderawasih mendapatkan jumlah siswa yang didapatkan dari kuesioner sangat memungkinkan orang tua mendaftarkan anaknya kesekolah ini. Dapat dilihat dari beberapa faktor yang dijadikan sebagai atribut dalam daftar kuesioner. Dengan algoritma Naïve Bayes, atribut yang paling tinggi sebagai faktor penentu orang tua mendaftarkan anaknya adalah atribut umur, yaitu yang berumur 31 sampai 50 tahun, kemudian atribut faktor karena tidak masuk negeri, atribut transport menggunakan motor, kemudian atribut jarak, serta atribut informasi yang didapat dari teman atau saudara. Dari faktor tersebut terlihat apa saja yang harus dilakukan untuk menaikkan jumlah siswa.

Kata Kunci: sekolah swasta, data mining, klasifikasi

Abstract

Competition in attracting parents students to send their children to private schools is very high indeed. Private schools must work hard in terms of getting students who will continue their education to the next level. Schools should have added value in the eyes of students parents. Of the three algorithm used, there is the highest accuracy value of 86,50% generated from the Naïve Bayes algorithm. In a sense, the prediction rate for Cenderawasih Middle School getting the number of students obtained from the questionnaire is very possible for parents to enroll their children in this school. Can be seen from several factors that serve as attributes in the questionnaire list. With the Naïve Bayes algorithm, the highest attribute as a determining factor for parents registering their children is the age attribute, that is received 31 to 50 years, then attribute factors for not entering the school country, transport attribute using a motor, then the distance attribute and attribute information obtained from friends or relatives. From this factors it can be seen what must be done to increase the number of students.

Pendahuluan

Saat ini banyak sekali bermunculan sekolah-sekolah swasta yang dikelola oleh sebuah yayasan untuk ikut berperan dalam mencerdaskan kehidupan bangsa. Dalam hal ini adalah membangun sumber daya manusia untuk mencapai kemampuan dalam bidang pendidikan. Dari tingkat paling bawah, yaitu taman kanak-kanak hingga tingkat menengah atas, tentunya pendidikan harus tetap diutamakan bagi setiap warga negara. Diharapkan dengan pendidikan yang bermutu dan berbiaya tidak mahal, setiap warga negara dapat terbantu dengan adanya sekolah swasta yang sangat banyak sekali keberadaannya.

Suatu promosi atau usaha untuk mengenalkan suatu sekolah, dapat dikatakan sangat membantu untuk dapat menarik minat para orang tua dalam menyekolahkan anaknya ketempat yang menurutnya lebih baik dibandingkan dengan sekolah lainnya. Promosi adalah alat atau media untuk melakukan komunikasi yang bersifat mempengaruhi calon konsumen agar tertarik membeli, menggunakan jasa,

atau sebuah produk yang dijual oleh sebuah instansi atau perusahaan (Gitosudarmono, Indriyo. 2000).

Adapun jenis dari promosi sangat banyak, seperti di media massa yang dapat dilakukan dengan mengiklankan produk yang akan dijual kepada masyarakat melalui koran, majalah, pamflet, brosur, banner dan sebagainya. Atau bahkan juga media promosi dapat dilakukan secara online seperti melalui media sosial yang dapat digunakan dengan cara mengunggah beberapa gambar pilihan dengan lengkap disertai dengan keterangan mengenai rincian kegiatan yang berada di tempat tersebut. Kemudian dapat ditambahkan ulasan lengkap terhadap beberapa program unggulan yang ditawarkan, rincian biaya, beberapa diskon yang akan diberikan dan sebagainya. Namun tentu saja ada kekurangan-kekurangan serta kelebihan yang terdapat dalam media promosi tersebut, sehingga ini membutuhkan suatu kajian yang lebih untuk mempelajari apa saja usaha yang paling tepat untuk diputuskan dan digunakan dalam kegiatan promosi sekolah. Selain kegiatan mengenalkan sekolah kepada masyarakat dilakukan juga usaha mendekati masyarakat dengan kegiatan sosial.

Pendididkan dasar adalah suatu tempat dimana seorang individu dengan status sebagai siswa mulai dari tingkat paling rendah hingga tingkat menengah atas, untuk menuntut ilmu dengan tujuan untuk membuat siswa tersebut menjadi lebih baik dalam kehidupannya. Dari mulai satu tahun di tingkat taman kanak-kanak enam tahun tingkat sekolah dasar, tiga tahun tingkat menengah pertama dan tiga tahun tingkat menengah atas.

Banyak usaha yayasan untuk mengenalkan sekolah yang berada di bawah naungannya dengan melakukan promosi ke sekolah-sekolah lain, menyebar pamflet, banner dan sebagainya, untuk tujuan meraih calon peserta didik agar tertarik mendaftar di sekolah tersebut. Akan tetapi, dalam melakukan promosi terkadang banyak kendala yang dialami yayasan, seperti kurangnya kebutuhan dana untuk menunjang promosi sekolah miliknya, kurangnya sumber daya manusia untuk menjalankan promosi pengenalan sekolah-sekolah milik yayasan, atau juga belum adanya keputusan yang tepat untuk mempromosikan sekolah-sekolah tersebut dalam penerimaan siswa baru.

Untuk mengatasi permasalahan promosi dalam penerimaan siswa baru tersebut, diperlukan sebuah metode yang digunakan untuk mencari faktor-faktor lain yang menentukan seseorang memutuskan untuk mendaftar kesekolah ini. Metode yang digunakan dalam penelitian ini menggunakan tiga metode algoritma, yakni algoritma C4.5, *Naïve Bayes* dan *k-Nearest Neighbor*.

Diharapkan dengan menggunakan tiga metode ini sebagai sistem pendukung keputusan untuk usaha menaikkan jumlah peserta didik baru dapat membantu pekerjaan tim promosi dan pimpinan dalam menentukan prioritas tindakan yang dilakukan serta langkah apa saja yang paling tepat.

A. Data *Mining*.

Menurut Turban et al mengatakan bahwa “data *mining* merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database* (Turban E., Aronson J.E., dkk (2003). Data *mining* adalah proses yang menggunakan Teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasikan informasi yang bermanfaat dan pengetahuan terkait dari berbagai *database* besar” (Kusrini & Luthfi, 2009).

Data *mining* juga diartikan sebagai suatu proses otomatis data-data yang sangat besar dan bertujuan untuk mendapatkan hubungan atau pola yang memberikan manfaat. Data *mining* juga merupakan suatu proses pendukung pengambil keputusan dimana mencari pola informasi dalam data. Pencarian ini dapat dilakukan oleh pengguna. Pencarian ini disebut *discovery*. *Discovery* merupakan proses pencarian dalam basis data dalam menemukan pola yang tersembunyi tanpa ide yang di dapatkan sebelumnya atau hipotesa tentang pola yang ada. Dengan kata lain aplikasi mengambil inisiatif untuk menemukan pola dalam data tanpa pengguna berfikir mengenai pertanyaan yang relevan terlebih dahulu,

Data *mining* bisa memberikan dampak negatif dan positif tergantung pada penggunaannya. Jika tidak memperhatikan etika penggunaan data, khususnya yang berhubungan dengan data pribadi pelanggan maka data *mining* bisa berdampak negatif. Misalnya klasterisasi pelanggan berdasarkan suku bangsa, agama, ras, golongan, usia maupun gender

bisa berujung pada masalah diskriminasi dan bisa merugikan suatu kelompok tertentu. Tetapi, ketika data *mining* digunakan untuk masalah medis yang harus membedakan *gender* dan usia tertentu, maka hal ini justru berefek positif. Misalnya, ada suatu jenis penyakit yang peluangnya lebih besar di derita oleh kaum wanita atau oleh kelompok usia tertentu, maka sudah seharusnya pihak medis melakukan penanganan secara berbeda.

B. Algoritma C4.5

Algoritma data *mining* C4.5 merupakan salah satu algoritma yang digunakan melakukan klasifikasi atau segmentasi atau pengelompokkan dan bersifat prediktif. Klasifikasi merupakan salah satu proses pada data *mining* yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar. Melibatkan konstruksi pohon keputusan, koleksi *node* keputusan, yang terhubung oleh cabang-cabang kemudian diperpanjang kebawah dari simpul akar sampai berakhir di *node* daun. Dimulai dari *node root*, yang oleh konvensi ditempatkan di bagian atas dari diagram pohon keputusan, atribut diuji pada *node* keputusan, dengan setiap hasil yang mungkin menghasilkan cabang. Setiap cabang kemudian mengarah ke *node* lain baik keputusan atau ke *node* daun untuk mengakhiri (Larose, 2005).

Algoritma C4.5 dan pohon keputusan (*decision tree*) merupakan dua model yang tidak dapat dipisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Dari akhir tahun 1970 sampai awal tahun 1980-an J.Ross Quinlan, melakukan pengembangan terhadap algoritma *decition tree* yakni ID3 (*Interative Dichotomisemiser*). Kemudian Quinlan juga menghadirkan algoritma C4.5 yang menjadi awal dari algoritma *supervised learning* yang terbaru. Di tahun 1984 sebuah kelompok statistik (L. Breimen, J. Fridman R. Olshen dan C. Stone) mempublikasikan *Classification and Regression Tree (CART)* yang menggambarkan generasi *binary decision tree.*

Terdapat beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Kusrini & Lutfi, 2009) yaitu :

1. Menyiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung *gain* tersebut, hitung dulu *entropy* yaitu dengan rumus:

Entropy (S) $= \sum_{i=1}^{n} -pi * \log 2\, pi$

C. Algoritma *Classifier Naïve Bayes*

Metode ini menggunakan teorema *Bayes* abad ke 18 (Suyanto, 2017). Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Menurut Wu dan Khumar bahwa *Naïve Bayes* merupakan metode klasifikasi popular dan masuk dalam sepuluh algoritma terbaik dalam data mining. *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data *training* (Wu, X., and Kumar, V. (2009).

Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary*, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sedapat mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi.

Naïve Bayes memiliki keunggulan yaitu kemudahan konstruksinya dan tidak membutuhkan parameter skema pengulangan yang kompleks sehingga mudah dalam membaca data dalam jumlah yang besar. Hal ini terjadi karena desain rancangan penuntunan klasifikasi terhadap data. Selain itu, metode ini dinyatakan sebagai algoritma yang mepunyai sifat *simplicity*, *elegance* dan *robustness*.

D. Algoritma *k-Nearest Neighbor*

Algoritma k-NN merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan tetangga terdekatnya dan kelas yang paling banyak

muncul yang akan menjadi kelas hasil klasifikasi (Witten, Frank, & Hall, 2011). Kelebihan dari *k-NN* adalah dapat digunakan untuk memecahkan permasalahan *multiclass* (Aburomman, Bin, & Reaz, 2015) namun, *k-NN* memiliki masalah untuk menemukan tetangga terdekat pada titik *query* dari *dataset* yang digunakan (Liaw, Wu, & Leou, 2010).

Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan sampel-sampel dari *training* data. Algoritma *k-Nearest Neighbor* menggunakan *Neighborhood Classification* sebagai nilai prediksi dari nilai *instance* yang baru. *K-Nearest Neighbor* adalah algoritma pengklasifikasian yang didasarkan pada analogi, yaitu membandingkan data uji dengan data pelatihan yang berada dekat dengan dan memiliki kemiripan dengan data uji tersebut (S. Tan, 2006). Kemiripan data uji dengan data pelatihan didasarkan pada jaraknya. Banyak persamaan yang dapat digunakan untuk menghitung jarak antara data uji dan data pelatihan.

Metode Penelitian

Penelitian ini menggunakan analisa data dari kuesioner terhadap 200 orang tua siswa yang hadir saat pelaksanaan acara promosi sekolah, dengan nama Gebyar SMP Cenderawasih. Para orang tua sekolah dasar tingkat akhir atau kelas enam, diundang untuk menghadiri acara tersebut. Undangan untuk acara disebar melalui sekolah-sekolah tingkat dasar yang ada di lingkungan terdekat dengan SMP Cenderawasih. Saat acara yang memang diadakan dengan tujuan para orang tua, guru dan siswa sekolah dasar untuk mengikuti berbagai perlombaan dan acara hiburan. Kemudian dilakukan kuesioner terhadap para orang tua yang hadir dengan masing-masing asal sekolah dasar negeri sebanyak 100 orang dan untuk sekolah dasar swasta sebanyak 100 orang. Dimana kuesioner terdiri dari 15 atribut, kemudian penulis juga menuliskan singkatan terhadap pertanyaan dengan tujuan agar lebih ringkas saat penulisan atribut dalam format excel. Atribut yang dimaksud antara lain : jenis sekolah (n/s), umur orang tua (umur), jenis kelamin (jk), jarak rumah anda ke sekolah sekarang (jarak), Pendidikan anda (Pendidikan), anak anda saat ini di sekolah dasar ikut ekstrakurikuler apa (ekskul), transportasi sehari-hari ke sekolah (transport), pendapatan keluarga setiap bulan (pendapatan), pengeluaran untuk Pendidikan setiap satu orang anak per bulan (alokasi), apa yang mempengaruhi jika anak anda masuk sekolah swasta (faktor), kualitas sekolah swasta (kualitas), tahu tentang SMP Cenderawasih (tahu), mengetahui info SMP Cenderawasih dari mana (info), yang anda ketahui tentang SMP Cenderawasih (kondisi) dan saran anda untuk SMP Cenderawasih (saran). Semua atribut yang akan dijawab oleh para kuesioner hanya di beri pilihan jawaban sebanyak satu saja. Teknis penjaringan para orang tua adalah dengan menanyakan asal sekolah anak mereka dan kemudian mereka mengisi kuesioner tersebut. Tentu saja hasil kuesioner yang diajukan diharapkan merupakan jawaban dari kondisi yang sebenarnya. Dengan kata lain, jawaban mereka seharusnya jujur.

Selanjutnya dilakukan pengujian dengan metode yang diusulkan terhadap *dataset* di atas, dengan kelas labelnya adalah jenis sekolah, hingga diperoleh suatu pola klasifikasi, hingga dapat dikatakan tepat dan tidak tepatnya orang tua dalam prediksi apakah anak mereka akan dimasukkan ke sekolah SMP Cenderawasih.

Pengumpulan data primer dalam penelitian ini menggunakan metode kuesioner terhadap para orang tua sekolah dasar dari negeri dan swasta. Juga memperhatikan dari data-data yang berhubungan dengan jumlah SMP Cenderawasih dalam lima tahun ke belakang, dimulai dari tahun pelajaran 2014-2015 hingga tahun pelajaran 2018-2019. Jumlah *dataset* sebanyak 200 record terdiri dari 15 atribut, parameter yang diuji sebelum dilakukan *preprocessing* antara lain Jenis sekolah, umur orang tua, jenis kelamin, jarak rumah anda ke sekolah sekarang, Pendidikan anda, anak anda saat di Sekolah Dasar ikut ekstrakurikuler apa, transportasi sehari-hari ke sekolah, pendapatan keluarga setiap bulan, pengeluaran untuk Pendidikan setiap satu orang perbulan, apa yang mempengaruhi jika anak anda masuk sekolah swasta, kualitas sekolah swasta, tahu tentang SMP Cenderawasih, mengetahui info SMP Cenderawasih dari mana, yang anda ketahui dari SMP Cenderawasih dan saran anda untuk SMP Cenderawasih. Pada pengumpulan data sekunder, menggunakan buku, jurnal, hasil publikasi dan informasi dari tata usaha sekolah.

Hasil dan Pembahasan

A. Hasil Penelitian.

Data-data penelitian ini tentang informasi diri dari para orang tua calon pendaftar di SMP Cenderawasih. Data-data tersebut diperoleh dari hasil angket atau kuesioner terhadap orang tua siswa sekolah dasar. Tujuan dari penelitian ini adalah untuk mengembangkan model yang telah terbentuk dengan algoritma C4.5, ditambah algoritma *Naïve Bayes* serta algoritma *k-Nearest Neighbor*. Data-data tersebut dianalisa dengan melakukan perbandingan menggunakan Algoritma C4.5, Algoritma *Naïve Bayes* dan Algoritma *k-Nearest Neighbor* yang semuanya akan di komparasi ke akuratannya. Diharapkan nantinya akan mendapatkan hasil yang lebih baik.

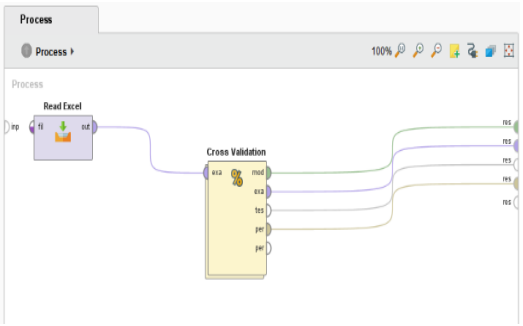
B. Evaluasi dan Validasi Metode.

Metode klasifikasi bisa di evaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, stabilitas dan *interpretabilitas* (Vercellis, 2009). Setelah data diolah maka dapat diuji tingkat akurasinya untuk melihat kinerja masing-masing metode. Penelitian ini bertujuan untuk melihat akurasi tingkat ketepatan calon siswa SMP Cenderawasih untuk dapat masuk ke sekolah tersebut, dengan harapan calon siswa akan lebih banyak mendaftar, kemudian melakukan perbandingan antara metode algortima C4.5, algoritma *Naïve Bayes*, serta algoritma *k-Nearest Neighbor*. Berdasarkan metode yang digunakan, metode mana tingkat akurasinya yang paling tinggi dengan menggunakan satu data *testing*.

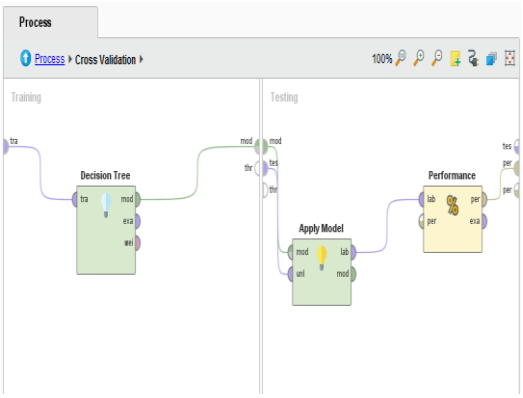
Melalui *dataset* yang terpilih di atas, maka selanjutnya dilakukan pengujian dan eksperimen dengan algoritma klasifikasi data *mining* melalui *framework* Rapidminer V9.3, dalam hal ini algoritma C4.5 dan algoritma *Naïve Bayes* serta algoritma *k-Nearest Neighbor*. Berikut ini adalah hasil dari setiap eksperimen dan hasil pengujian yang telah dilakukan :

1. Algoritma C4.5 (*decition tree*)

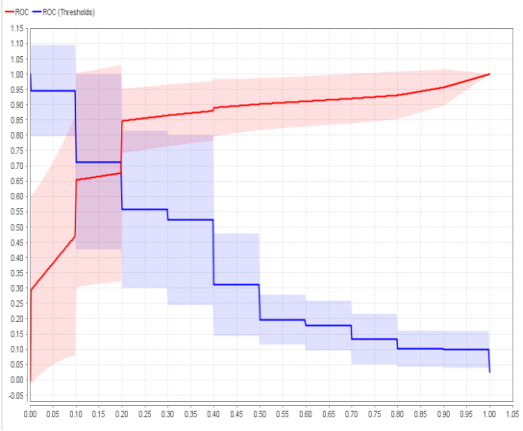
Hasil pengujian yang dilakukan melalui model C4.5 menghasilkan *Confusion Matrix*, *Accuracy* sebesar 79,50%, *Precision* sebesar 78,64% dan *Sensitivity* atau *Recall* 81,00% seperti terlihat pada gambar di 1, 2 dan 3.



Gambar 1. Desain algoritma *C4.5*



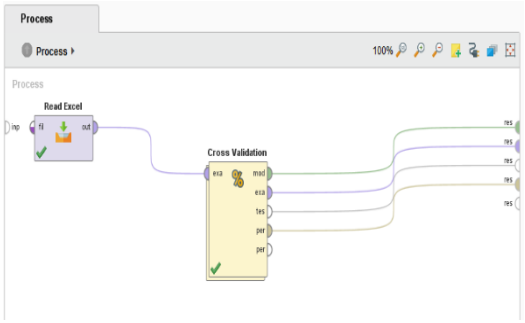
Gambar 2. *Validation* model algoritma C4.5



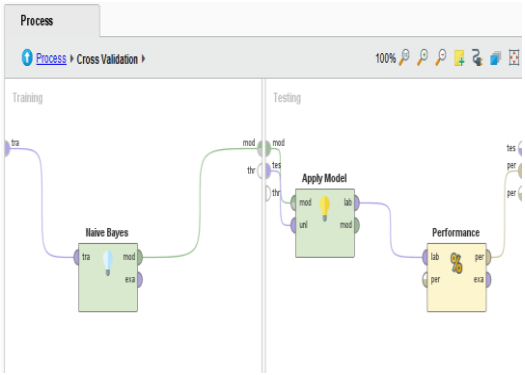
Gambar 3. *AUC* model algoritma C4.5

2. Algoritma *Naïve Bayes*

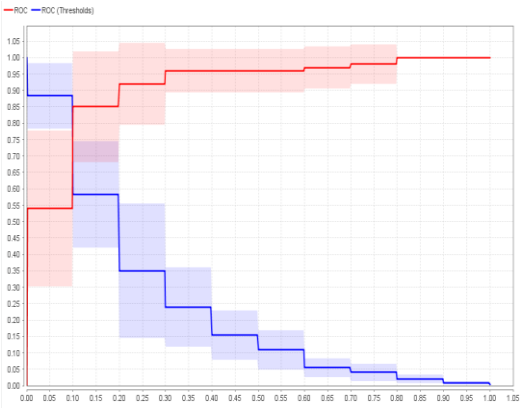
Selanjutnya hasil pengujian yang dilakukan melalui model *Naïve Bayes* menghasilkan *Confusion Matrix*, yaitu *accuracy* sebesar 86,50%, *Precision* sebesar 86,14%, dan *Sensitivity* atau *Recall* sebesar 87,00% seperti terlihat pada gambar 4, 5, dan 6.



Gambar 4. Desain *Naïve Bayes*



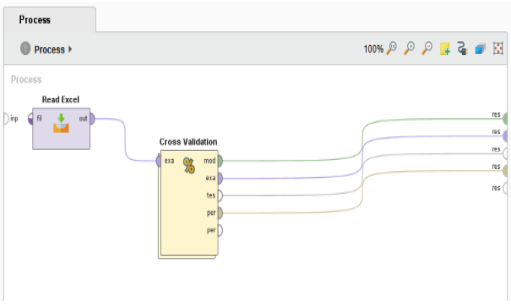
Gambar 5. *Validation Naïve Bayes*



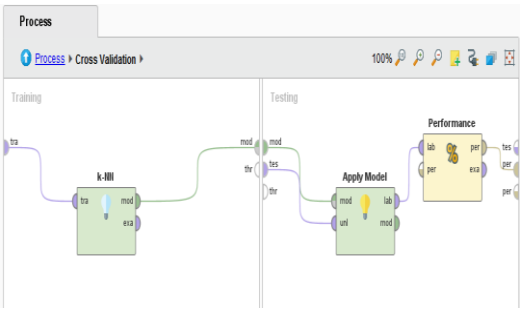
Gambar 6. *AUC* model *Naïve Bayes*

3. Algoritma *k-Nearest Neighbor*

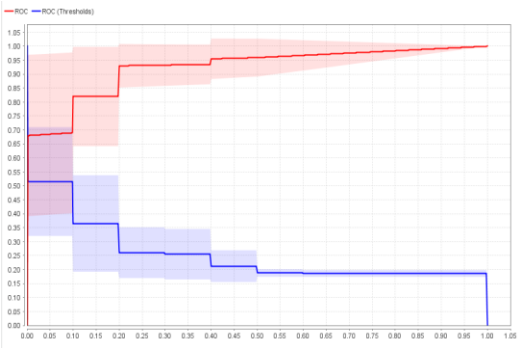
Hasil pengujian yang dilakukan melalui model *k-Nearest Neighbor* menghasilkan *Confusion Matrix*, yaitu *Accuracy* sebesar 80,00%, *Precision* sebesar 73,44%, dan *Sensitivity* atau *Recall* sebesar 94,00% seperti terlihat pada gambar 7, 8, dan 9.



Gambar 7. Desain *k-Nearest Neighbor*



Gambar 8. *Validation k-Nearest Neighbor*



Gambar 9. *AUC k-Nearest Neighbor*

Dari hasil eksperimen yang dilakukan dengan beberapa model antara lain algoritma C4.5, model algoritma *Naïve Bayes* dan model *k-nearest neighbor* maka dapat dilihat pada tabel 1 di bawah ini :

Criteria	*C45*	*Naïve Bayes*	*k-Nearest Neighbor*
Accuracy	79,50%	86,50%	80,00%
Precission	78,64%	86,14%	73,44%
Recall	81,00%	87,00%	94,00%

Tabel 1. *Accuracy* Algoritma Klasifikasi Data *Mining*

Di lihat dari tabel 1, bahwa *accuracy* dari algoritma *Naïve Bayes* merupakan *accuracy* yang paling tinggi yaitu 86,50% dibandingkan dengan algoritma C4.5 yang

sebesar 79,50% dan algoritma *k-Nearest Neighbor* sebesar 80,00%.

Kesimpulan

Hasil eksperimen terhadap *accuracy* dengan algoritma C4.5 adalah sebesar 79,50% dan kemudian dengan algoritma *Naïve Bayes* di dapat nilai *accuracy* sebesar 86,50% serta juga dengan perhitungan *accuracy* terhadap algoritma *k-nearest neighbor* di dapat nilai sebesar 80,00%.

Dari ketiga algoritma yang digunakan terdapat nilai *accuracy* yang tertinggi adalah sebesar 86,50% yang dihasilkan dari algoritma *Naïve Bayes*. Dalam artian tingkat prediksi untuk sekolah SMP Cenderawasih mendapatkan jumlah siswa yang di dapat dari data-data kuesioner, sangat memungkinkan sekali para orang tua siswa akan mendaftarkan anaknya ke sekolah ini. Dapat dilihat dari beberapa faktor yang dijadikan sebagai atribut dalam kuesioner yang telah di tanyakan dan dijawab oleh para orang tua siswa. Dari perhitungan dengan metode algoritma *Naïve Bayes*, diketahui atribut yang paling tinggi untuk faktor penentu orang tua siswa mendaftarkan anaknya ke sekolah SMP Cenderawasih adalah atribut umur, yang berumur 31 sampai 50 tahun, kemudian atribut faktor, karena tidak masuk negeri, atribut transportasi, menggunakan motor, kemudian atribut jarak, dimana rumah mereka diatas 1 km, serta atribut informasi, dimana informasi tentang SMP Cenderawasih di dapat dari teman atau saudara. Dari faktor-faktor tersebut terlihat apa saja yang harus dilakukan untuk menaikkan jumlah siswa.

Referensi

Gitosudarmono, Indriyo. (2000). Manajemen Pemasaran. Edisi II, BPFE, Yogyakarta.

Turban E., Aronson J.E., dkk (2003). Decision Support System and Intelligent System (Sistem Pendukung Keputusan dan Sistem Cerdas), Andi Offset, Yogyakarta.

Kusrini,, Luthfi Taufiq Emha (2009). Algoritma Data Mining, Penerbit Andi Offset, Yogyakarta.

Larose, Daniel T, (2005). Discovering Knowledge in Data : an Introduction to Data Mining, John Willey & Sons Inc.

Suyanto. (2017). Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung : Informatika Bandung

Wu, X., and Kumar, V. (2009). The Top Ten Algorithms in Data Mining. Boca Raton, London, New York: Taylor & Francis Group, LLC.

Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining : Practical Machine Learning and Tools. Burlington: Morgan Kaufmann Publisher.

Aburomman, A. A., Bin, M., & Reaz, I. (2015). A Novel SVM-kNN-PSO ensemble method for intrusion detection system. Applied soft Computing Journal, 1-13. https://doi.org/10.1016/j.asoc.2015.10.011.

Liaw, Y., Wu, C., & Leou, M. (2010). Fast k-nearest neighbor search using modified principal axis search tree. Digital Signal Processing, 20(5), 1494-1501. https://doi.org/10.1016/j.dsp.2010.01.009.

Tan, Pang Ning. Michael. Steinbach, and Vipin. Kumar (2006). Introduction To Data Mining. 1st Penyunt. Boston: Pearson Addison Wesley.

Vercellis, Carlo. (2009). Business Intelligence: Data Mining and Optimization for Decision Making. United Kingdom: John Willey & Son.