Jurnal Rekayasa Teknologi Nusa Putra. Vol. 5, No. 2, Februari 2019: Hal 15 - 19

PEMILIHAN JURUSAN DENGAN METODE K-NEAREST NEIGHBOR UNTUK CALON SISWA BARU

[1]Sudin Saepudin, [2]Muhamad Muslih, [3]Sihabudin
[1,2,3]Program Studi Sistem Informasi
[1,2,3]Universitas Nusa Putra
[1,2,3]Jl. Raya Cibolang Kaler No.21 Kab. Sukabumi
e-mail: [1]sudin.saepudin@nusaputra.ac.id,[2]muhamad.muslih@nusaputra.ac.id, sihabudin@nusaputra.ac.id

Korespondensi : [1]sudin.saepudin@nusaputra.ac.id

ABSTRAK

Kebutuhan akan analisis mengenai prediksi di SMK YASPIM, menjadi alasan dilakukannya penelitian terhadap prediksi pemilihan jurusan siswa, Menentukan prediksi pemilihan jurusan dalam jumlah besar tidak mungkin dilakukan secara manual karena membutuhkan waktu yang cukup lama. Untuk itu dibutuhkan sebuah algoritma yang dapat memecahkan prediksi secara otomatis menggunakan komputer. Dalam memprediksi digunakan sistem pendukung keputusan berupa aplikasi weka dan dalam menganalisis menggunakan Metode K-Nearest Neighbor karena memiliki prinsip kerja mencari jarak terpendek antara data yang akan dievaluasi dengan K terdekat dalam data pelatihannya. Berdasarkan pengujian hasil prediksi dengan menggunakan Algoritma K-Nearest Neighbor yang pengujiannya dilakukan secara real dan menggunakan sistem pendukung keputusan menghasilkan data yang baiknya.Hasil uji tersebut menyatakan system cukup layak prediksi pemilihan jurusan meskipun sistem belum menghasilkan tingkat akurasi yang maksimal. Untukmenghasilkan nilai validasi yang maksimal membutuhkan data yang seimbang antara kasuskeluar dan aktif. Hasil pengujian menggunakan weka sistem pendukung keputusan dandibandingkan secara manual yang menggunakan metode Algoritma K-Nearest Neighbor dengan4 variabel yaitu nilai rata-rata ujian nasional, rata-rata nilai raport, minat dan rata-rata nilai hasil test adalah mendapatkan kesesuaian data yang diharapkan smk yaspim.

Kata kunci : Metode Algoritma Data Mining, K-Nearest Neighbor.

ABSTRACT

The need for analysis regarding predictions at SMK YASPIM is the reason for conducting research on prediction of student majors selection. Determining predictions for choosing a large number of majors is impossible to do manually because it takes a long time. For that we need an algorithm that can solve predictions automatically using a computer. In predicting the use of a decision support system in the form of a weka application and in analyzing it using the K-Nearest Neighbor method because it has a working principle of finding the shortest distance between the data to be evaluated and the nearest K in the training data. Based on testing the prediction results using the K-Nearest Neighbor Algorithm, the tests are carried out in real terms and using a decision support system to produce good data. The test results show that the system is quite feasible for predicting the selection of majors even though the system has not produced the maximum level of accuracy. To produce the maximum validation value requires balanced data between outgoing and active cases. The test results using the Weka decision support system and compared manually using the K-Nearest Neighbor Algorithm method with 4 variables, namely the average value of the national exam, the average value of report cards, interest and the average value of the test results is to get the expected data suitability SMP Yaspim .

Keywords : Data Mining Algorithm Method, K-Nearest Neighbor.

I. PENDAHULUAN

Dalam dunia pendidikan kita mengenal yang namanya siswa atau murid setiap tahun sekali banyak siswa yang masuk dan keluar sekolah di dalam penerimaan siswa baru, tersebut beramai-ramai mendaftarkan diri menjadi calon siswa sekolah setiap tahunya siswa masuk kesekolah menjadi calon siswa mencapai ratusan bahkan ribuan. SMK YASPIM merupakan sekolah Komputer yang mempunyai beberapah jurusan yaitu Teknik Komputer dan Jaringan (TKJ), Rekayasa Perangkat Lunak (RPL), Teknik Sepeda Motor (TSM) maka dari itu harus ada prediksi untuk masuk kedalam tiga jurusan tersebut.

Setiap pekerjaan dapat di selesaikan dengan menggunakan teknologi baik komputerisasi atau dengan metode manual, SMK YASPIM harus bisa memprediksi dari jumlah siswa yang mendaftar untuk masuk kedalam tiga jurusan, karena tidak semua siswa yang daftar ke SMK YASPIM tersebut memilhih sesuai dengan keinginan calon siswa. SMK YASPIM perlu memprediksi siswa yang masuk atau memilih jurusan supaya sekolah dapat mempersiapkan segala keperluan mulai dari kelas SDM guru dan lain-lain, maka dari itu sebelum ada hasil tes sekolah harus dapat memprediksi jumlah siswa yang daftar dan berapa kelas dari masing-masing jurusan serta harus memberikan data akurasi supaya ada pertimbangan dari sebelumnya maupun yang akan dating.

Algoritma nearest neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. KNearest neighbor termasuk algoritma supervised learning, dimana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada nearest neighbor. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi, metode ini lebih efektif di data traning yang besar dapat menghasilkan data yang lebih akurat tetapi metode nearest neighbor perlu menentukan nilai k yang paling optimal yang menyatakan jumlah tetangga terdekat dan biaya komputasi cukup tinggi karena perhitungan jarak harus dilakukan pada setiap query instance bersama-sama dengan seluruh instan dari traning sample.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

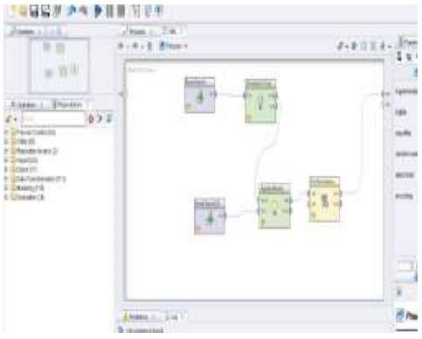
Prinsip kerja k-Nearest Neighbour (K-NN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga (Neighbour) terdekatnya dalam data pelatihan [1]. Algoritma k-NN adalah suatu metode yang menggunakan algoritma supervised [2].

NBC adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. NBC didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa decision tree dan neural network. Naïve Bayes merupakan salah satu metoda machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [3].

Algoritma pengklasifikasian data sederhana dimana penghitungan jarak terpendek dijadikan ukuran untuk mengklasifikasikan suatu kasus baru berdasarkan ukuran kemiripan. Algoritma k-NN tergolong dalam algoritma supervised yaitu proses pembentukan algoritma diperoleh melalui proses pembelajaran (learning) pada record-record lama yang sudah terklasifikasi dan hasil pembelajaran tersebut dipakai untuk mengklasifikasikan record baru dengan output yang belum diketahui [4].

2.2 Rapid Miner 7

Rapid Miner adalah aplikasi data mining yang tidak perlu dipertanyakan lagi dan berbasis sistem open-source dunia yang terkemuka dan ternama. Tersedia sebagai aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining untuk integrasi ke dalam produk sendiri.



Gambar 1. Tampilan Rapid Miner

III. METODOLOGI PENELITIAN

Metode Pemilihan Sampel, Sampel pada penelitian ini adalah calon siswa SMK YASPIM Gegerbitung Sukabumi.dari tiga tahun dengan sekarang berikut adalah data sempel dari tahun ketahun.

3.1 Perancangan Nearest Neighbor

Perancangan pendekatan mencari kasus dengan menghitung kedekatan antara kasus baru dan kasus lama,yaitu berdasarkan pada pencocokan bobot dari jumlah fitur yang ada.

Adapun rumus yang di gunakan dapat dilihat pada persamaan berikut:

1. Input nilai kriteria masing-masing model
2. Input bobot masing-masing kriteria
3. Menghitung normalisasi dari bobot

$$NK=\frac{\sum_{i=i}^{n}(SBK)xBBT\,\%}{n}$$

$$\text{Nilai Akhir} =\frac{\sum^{n} NK}{n}$$

Dimana :SBK : Kriteria
BBT : Bobot Kriteria
(Pembulatan dari nilai Kriteria)
NK :Nilai Kriteria

Pemberian Nilai :

- Pemberian Pembobot pada nilai UN
- Pemberian Pembobotan pada Nilai Raport
- Pembobotan pada nilai Test d. Pembobotan minat siswa

3.2 Teknik Analisis Data

Analisis data menggunakan Data Kuantitatif berupa kaidah-kaidah matematika terhadap data. Penerapan algoritma data mining nearest neighbor menggunakan bantuan software rapidminer dengan menggunakan fasilitas yang disediakan pada toolbox rapidminer. Langkah-langkah penelitian yang dilakukan ini menggunakan model Cross-Standard Industry for Data Mining (CRISP-DM) yang terdiri dari 6 fase, yaitu:

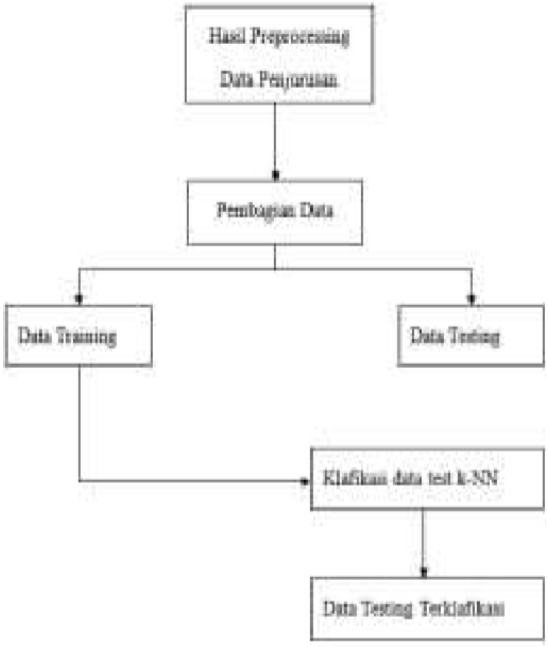
1. Business/Research Understanding Phase Berdasarkan data yang diterima dari tim seleksi penerima siswa baru, ternyata banyak siswa yang daftar memilih jurusan yang diinginkankan tetapi sekolah harus bisa memprediksi siwa yang dapat masuk kedalam tiga jurusan karena setiap jurusan mempunyai batas siswa oleh dengan menggunakan metode nearest neighbor di bantu oleh aplikasi rapidminer dapat memberikan akurasi yang lebih akurat.
2. Data Understanding Phase (Fase Pemahaman Data) Untuk pemilihan jurusan, didapat data dari SMK YASPIM Sukabumi sebanyak 250 data siswa , terdiri dari 4 Atribut-atribut yang menjadi parameter terlihat pada tabel 1 yaitu:

Tabel 1. Atribut penelitian

No	Atribut	Nilai
1	Nilai Ujian Nasional	30%
2	Nilai Raport	25%
3	Nilai Test	35%
4	Minat	10%
	Jumlah	100%

3. Data Preparation Phase (Fase Pengolahan Data) Data yang diperoleh untuk penelitian ini sebanyak 250 record siswa . Baik yang memilh sesuai keinginan ataupun kejurusan lain . Untuk mendapatkan data yang berkualitas, ada beberapa teknik preprocessing yang digunakan, yaitu:
   - Data validation, untuk mengidentifikasikan dan menghapus data yang tidak konsisten dan data yang tidak lengkap (missing value)
   - Data Integration and Transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini kategorikal, untuk model nearest neighbor data ditransformasikan ke dalam angka menggunakan software RapidMiner.
   - Data size reduction and dicretization, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informative. Didalam data training yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi menggunakan software RapidMiner.
4. Modeling Phase (Fase Pemodelan) Tahap ini juga disebut tahap learning karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini, pembuatan model menggunakan algoritma neural network.
5. Evaluation Phase (Fase Evaluasi) Pada tahap ini dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang

akurat. Evaluasi dan validasi menggunakan metode Confision Matrix dan kurva ROC.Deployment Phase (Fase Penyebaran) Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini juga diterapkan model yang paling akurat ke sekolah untuk pemilihan penerima beasiswa.



Gambar 2. Teknik Analisis Data

IV. HASIL DAN PEMBAHASAN

Pembuatan model algoritma Nearest Neighbor diawali dengan pembacaan file data (Read Excel). Data training disimpan dalam satu file Excel 2003. Lalu data di transform ke bentuk numeric, Langkah selanjutnya adalah penentuan model algoritmanya.

Gambar 3. Konfigirasi data dalam rapid miner

1. Evaluasi dan Validasi Dalam penulisan ini metode yang digunakan, yaitu algoritma , kemudian dilakukan pemilihan keduanya dan mengukur metode mana yang paling akurat. Metode klasifikasi bisa dievaluasi berdasarkan beberapa kriteria seperti tingkat akurasi, kecepatan, kehandalan, skalabilitas. (Vecellis, 2009) Penelitian ini bertujuan untuk mengukur tingkat akurasi dua algoritma dalam memilih penerima beasiswa, kemudian menganalisa akurasi dengan membandingkan kedua metode tersebut. Adapun tahapan-tahapan dalam evaluasi dan validasi adalah sebagai berikut:
   a. Pengujian Model Model yang telah dibentuk diuji tingkat akurasinya dengan memasukan data uji yang berasal dari data training. Karena data yang didapat dalam penelitian ini setelah proses preprocessing hanya 37 data maka digunakan metode cross validation untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk algoritma neural Nearest Neighbor sebesar 92.8%.

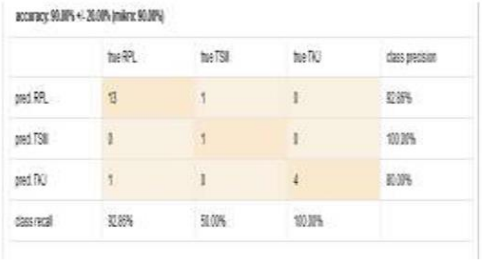
Gambar 4. Model Validasi

Selain itu dalam penelitian ini diuji juga dengan menggunakan confution matrix dan kurva ROC.

4.1 Confusion Matrix

Tabel IV.4 diketahui dari 37 data, 12 diklasifikasikan ya sesuai dengan prediksi yang dilakukan dengan metode Nearest Neighbor, kemudian 1 data diprediksi ya tetapi ternyata hasilnya prediksi tidak, 24 data remark tidak diprediksi sesuai, dan 0 data diprediksi tidak ternyata hasil prediksinya ya.

Tabel 2 Model Confusion Matrix Nearest Neighbor

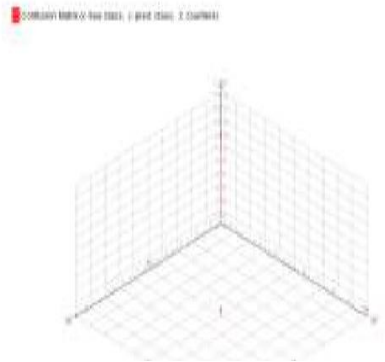


	true RPL	true TSW	true TKJ	class precision
pred. RPL	13	1	0	92.86%
pred. TSW	0	1	0	100.00%
pred. TKJ	1	0	4	80.00%
class recall	92.86%	50.00%	100.00%	

Berdasarkan Tabel 2 tersebut menunjukan bahwa, tingkat akurasi dengan menggunakan algoritma Nearest Neighbor adalah sebesar 92.86%.

4.2 Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Gambar 5 adalah kurva ROC untuk Nearest Neighbor



Gambar 5. Kurva RoC yang dihasilkan

4.3 Implikasi Penelitian

Dari hasil evaluasi ternyata algoritma Nearest Neighbor terbukti akurat untuk digunakan pada data mining pemilihan jurusn. Dengan demikian algoritma Nearest Neighbor dapat memberikan pemecahan untuk permasalahan penentuan kelayakan siswa dalam masuk kedalam jurusan. Untuk mendukung pengambilan keputusan dan pengembangan sistem informasi manajemen stratejik, model ini dapat diterapkan pada SMK YASPIM Sukabumi menggunakan rapidminer. Penelitian ini dapat dikembangkan pada unit lain, penelitian ini juga dapat dikembangkan dengan allgoritma yang lain untuk bisa membandingkan dengan algoritma yang sudah digunakan.

4.4 Aspek Sistem

1. Dikarenakan aplikasi prediksi ini adalah program yang memerlukan perangkat yang baik, pihak SMK harus menyediakan infrastruktur yang lebih baik, sehingga dalam penerapannya akan berjalan dengan maksimal.
2. Pihak lembaga harus meningkatkan kualitas sistem, meng-upgrade perangkat keras terutama komputer yang digunakan dalam program ini. Peningkatan memory, processor, media penyimpanan dan software aplikasi, diharapkan dapat meningkatkan unjuk kerja dan stabilitas sistem.
3. Meningkatkan dari sisi keamanan pada program ini yang diterapkan.
4. Diperlukan sistem perawatan yang baik terhadap infrastruktur pendukung. Dengan perawatan infrastruktur yang baik, masalah yang timbul pada aplikasi ini berjalan lancar.

4.5 Aspek Manajerial

1. Pihak SMK meningkatkan strategi untuk memotifasi calon siswa dalam bersaing menentukan jurusan .
1. Pihak lembaga harus meningkatkan kualitas informasi, kualitas sistem dan kualitas pelayanan, sehingga informasi mengenai beasiswa dapat segera diketahui.

V. PENUTUP

5.1 Kesimpulan

Dari uraian yang telah dijelaskan pada bab-bab sebelumnya, dapat disimpulkan hal-hal sebagai berikut: Dengan menggunakan klasifikasidata mining, algoritma yang berbasis prediksi penerapan algoritma k-nearest neighbor menghasilkan keputusan yang akurat dengan teknik konvensional untuk memprediksi pemilihan jurusan, Memberikan kemudahan kepada SMK YASPIM terutama memprediksi pemilihan jurusan, Metode algoritma knearest neighborini bisa memproses data yang lebih banyak, k-nearest neighbor mampu memprediksi masalah dan akurasi sangat baik dalam perbandingan kedekatan kasus baru dan kasus lama dan k-nearestn neighborharus menentukan nilai k dan bobot dari tiap-tiap variable.

5.2 Saran

Sebaikya kedepan digunakan data yang lebih banyak dan komplek, hal ini agar lebih optimal dalam proses akurasi metode yang dipilih.

DAFTAR PUSTAKA

[1] R. Yepriyanto et.al, "Sistem Diagnosa Kesuburan Sperma Dengan Metode K-Nearest Neighbor (K-NN)", Jurnal Ilmiah SINUS, 2014.

[2] Nursalim et. al, "Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma K-Nearest Neighbor", Pascasarjana Teknik Informatika, Universitas Dian Nuswantoro, Jurnal Teknologi Informasi, Vol. 10 No. 1, April 2014

[3] S.Listyaningrum, "Penerapan Data Mining untuk Analisis Karakteristik DPT Non-Participate sebagai Prediksi Partisipan Pemilu dengan Menggunakan Metode Naive Bayes Classifier", Program Studi Teknik Informatika, Fakultas Ilmu KomputeR, Universitas Dian Nuswantoro Semarang, 2015

[4] A. Nugroho, "Implementasi Algoritma K-Nearest Neighbor dalam Memprediksi Potensi Calon Kreditur Di Ksp Galih Manunggal", Jurnal Ilmiah DASI Vol. 17 No. 2 Juni 2016