

IMPROVE

Jurnal Ilmiah Manajemen Informatika
Jurnal Ilmiah Manajemen Informatika

PENERAPAN ALGORITMA K-NEAREST NEIGHBORS UNTUK
MENENTUKAN POLA PENERIMAAN MAHASISWA BARU

Mubassiran, M. Ibnu Choldun R.

PENATAAN DESA WISATA DI CIHANJUANG

Maniah, Shiyami Milwandhari, M. Ibnu Choldun R.

SISTEM INFORMASI WEIGHING BERBASIS WEB
(STUDI KASUS: PT. KALBE MORINAGA INDONESIA)

Fahriza Suryanto, Mubassiran, Virdiandry Putratama

RANCANG BANGUN SISTEM INFORMASI PENGADAAN MATERIAL
PENUGASAN (STUDI KASUS: PT. PLN (PEERSERO) PUSHARLIS UWP III)

Lina Karlina, Virdiandry Putratama, Mubassiran

RANCANG BANGUN APLIKASI E-TICKETING UNTUK MENINGKATKAN
PELAYANAN BAGI PENGUNJUNG OBJEK WISATA

Mochammad Wildan Syakuro, Mubassiran, Sari Armiati

ANALISIS DAN PERANCANGAN SISTEM INFORMASI STORAGE
CONTAINER PADA CONTAINER YARD (CY) DIVISI TPS (TEMPAT
PENIMBUNAN SEMENTARA) PT. INDONESIAN AIR & MARINE SUPPLY

Amri Yanuar, Dian Prayusman, Febriani Sulistianingsih

PERANCANGAN SMART SYSTEM SERVICE DI DESA BERBASIS SMS
GATEWAY

Supono, Sari Armiati

1

8

13

19

24

30

38

IMPROVE

Jurnal Ilmiah Manajemen Informatika

Volume 11 Nomor 2 Tahun 2019

Penerapan Algoritma K-Nearest Neighbors untuk Menentukan Pola Penerimaan Mahasiswa Baru	1
Mubassiran, M. Ibnu Choldun R.	
Penataan Desa Wisata di Cihanjuang	8
Maniah, Shiyami Milwandhari, M. Ibnu Choldun R.	
Sistem Informasi Weighing Berbasis Web (Studi Kasus: PT. Kalbe Morinaga Indonesia)	13
Fahriza Suryanto, Mubassiran, Virdiandry Putratama	
Rancang Bangun Sistem Informasi Pengadaan Material Penugasan (Studi Kasus: PT. PLN (PERSERO) PUSHARLIS UWP III	19
Lina Karlina, Virdiandry Putratama, Mubassiran	
Rancang Bangun Aplikasi e-Ticketing untuk Meningkatkan Pelayanan Bagi Pengunjung Objek Wisata	24
Mochammad Wildan Syakuro, Mubassiran , Sari Armiati	
Analisis dan Perancangan Sistem Informasi Storage Container pada Container Yard (CY) Divisi TPS (Tempat Penimbunan Sementara) PT. Indonesian Air & Marine Supply	30
Amri Yanuar, Dian Prayusman, Febriani Sulistianingsih	
Perancangan Smart System Service di Desa Berbasis SMS Gateway	38
Supono, Sari Armiati	



Politeknik Pos Indonesia

ISSN: 1979 - 8342

IMPROVE

INFORMATICS-MANAGEMENT-PROFESSIONAL-VOCATIONAL-ENTERPRISE

Publisher:

Jurusan Manajemen Informatika -
Politeknik Pos Indonesia
ISSN 1979-8342

Editorial Director

Virdiandry Putratama, S.T.

Advisory Board

Ari Yanuar, S.T., M.T.
Sari Armiami, S.T., M.T.
Saepudin Nirwan, S.Kom., M.Kom.

Editor in Chief

Maniah, S.Kom., M.T.

Editorial Board

Shiyami Milwandhari, S.Kom., M.T.
Supono, S.T., M.T.
Mubassiran, S.Si., M.T.
Ibnu Choldun, S.T., M.T.

Editorial Address

Jurusan Manajemen Informatika -
Politeknik Pos Indonesia
Jl Sariasih 54 Bandung
Telp, 022-2009570

Salam Manajemen Informatika,

Majalah ini merupakan sarana publikasi ilmiah, yang merupakan hasil kolaborasi antara mahasiswa dan dosen-dosen di jurusan Manajemen Informatika serta dosen dari luar Politeknik Pos Indonesia.

Pada edisi kali ini naskah tulisan diperoleh dari hasil penelitian dosen dan mahasiswa jurusan Logistik Bisnis, dan dosen dan mahasiswa jurusan Manajemen Informatika Politeknik Pos Indonesia.

Untuk itu kami mengucapkan terima kasih kepada para penulis dan juri yang telah meluangkan waktunya dalam menjamin mutu publikasi ilmiah ini. Semoga media ini dapat menjadi salah satu cara di jurusan Manajemen Informatika, menuju arah yang lebih baik lagi pada masa-masa yang akan datang, Aamiin YRA.

Redaksi

PENERAPAN ALGORITMA K-NEAREST NEIGHBORS UNTUK MENENTUKAN POLA PENERIMAAN MAHASISWA BARU (Studi Kasus: Politeknik Pos Indonesia)

¹Mubassiran, ²M. Ibnu Choldun

^{1,2}Program Studi D III Manajemen Informatika Politeknik Pos Indonesia

¹mubassiran@poltekpos.ac.id, ²ibnuholdun@poltekpos.ac.id

Abstrak

Penerimaan mahasiswa baru (PMB) adalah proses penyaringan calon mahasiswa yang diterima pada suatu perguruan tinggi. Mengidentifikasi pola dari PMB dapat memberikan informasi yang bermanfaat baik kepada perguruan tinggi dalam hal ini program studi atau calon mahasiswa yang mendaftar pada suatu program studi. Penentuan pola tersebut dapat dilakukan dengan model klasifikasi, model klasifikasi dibuat dengan cara menganalisis training data, model yang dihasilkan nantinya dapat digunakan untuk memprediksi kelas dari *unknown* data. Model klasifikasi dapat digambarkan dalam berbagai bentuk, salah satunya adalah dengan menggunakan k-Nearest Neighbors. Dalam penelitian ini akan dibahas model klasifikasi menggunakan k-Nearest Neighbors untuk penentuan pola dari data PMB dengan mengacu pada parameter atribut yang digunakan pada saat calon mahasiswa tersebut mendaftar dan melaksanakan ujian masuk.

Kata kunci: Algoritma K-Nearest Neighbors, Penerimaan Mahasiswa Baru, Data mining.

I. PENDAHULUAN

1.1 Latar Belakang

Suatu hal penting yang seharusnya dilakukan perguruan tinggi pada saat melakukan PMB, yaitu dengan mengidentifikasi pola dari data PMB yang sudah dilaksanakan, dengan melakukan klasifikasi dari parameter atribut yang ditentukan. Informasi yang dapat diterima program studi dari penentuan pola tersebut dapat dijadikan dasar penentuan kebijakan sistem PMB yang akan datang. Informasi yang dihasilkan juga bermanfaat bagi calon mahasiswa yang akan mendaftar pada suatu program studi, informasi tersebut dapat dijadikan dasar untuk pemilihan program studi dengan informasi kriteria penilaian suatu program studi terhadap calon mahasiswanya.

Klasifikasi adalah suatu fungsionalitas yang akan menghasilkan model yang mampu memprediksi kelas atau kategori dari objek-objek. Dalam permasalahan ini klasifikasi dapat digunakan oleh suatu program studi untuk menentukan atau mengidentifikasi pola dari data PMB yang sudah dilaksanakan. Pada kasus ini model klasifikasi dibuat untuk mengidentifikasi pola data untuk kelas status “melakukan Her-Registrasi” atau “tidak Melakukan Her-Registrasi”, dari hasil penentuan pola training data. Pola atau model dari training data

tersebut selanjutnya diuji dengan menggunakan test dataset. Singkatnya Model klasifikasi dibuat dengan cara menganalisis training data (terdiri dari variable variable yang kelasnya sudah diketahui). Model yang dihasilkan kemudian akan digunakan untuk memprediksi kelas dari unknown data (variable-variabel yang kelasnya belum diketahui), test dataset digunakan untuk pengujian dari model yang telah didapatkan pada training data. Selama ini Poltekpos tidak mengetahui profil mahasiswa seperti apa yang melakukan Her Registrasi di kampus Politeknik Pos Indonesia. Sehingga strategi penerimaan mahasiswa baru masih menggunakan teknik konvensional. Model klasifikasi yang digunakan dalam kasus ini adalah k-Nearest Neighbors. Perangkat lunak bantu yang digunakan untuk implementasi adalah RapidMiner. Diharapkan dengan dilakukannya model klasifikasi ini institusi ataupun prodi akan menemukan pola ataupun profil mahasiswa dari data PMB yang sudah dilaksanakan, sebagai dasar untuk penentuan kebijakan bagi PMB yang akan datang.

1.2 Perumusan Masalah

Berdasarkan latar belakang penelitian maka diperoleh identifikasi sebagai berikut:

Berdasarkan dari latar belakang di atas dapat dirumuskan : “Bagaimana pemodelan Klasifikasi

dari data Penerimaan Mahasiswa Baru dengan menggunakan algoritma k-Nearest Neighbors dan bantuan aplikasi RapidMiner untuk mendapatkan profil Mahasiswa Baru sebagai dasar untuk penentuan kebijakan bagi PMB yang akan datang di Kampus Politeknik Pos Indonesia.

1.3 Tujuan

Berdasarkan rumusan masalah di atas maka tujuan penelitian adalah menemukan model atau pola dari data Penerimaan Mahasiswa Baru dengan menggunakan algoritma k-Nearest Neighbors dan bantuan aplikasi RapidMiner untuk mendapatkan profil Mahasiswa Baru sebagai dasar untuk penentuan kebijakan bagi PMB yang akan datang di Kampus Politeknik Pos Indonesia.

1.4 Metodologi Penelitian

Prosedur Penelitian Penelitian ini pertama diawali dengan tahap identifikasi dan pengumpulan data, pencarian literatur, kemudian dilanjutkan dengan tahap persiapan dan pemilihan data, selanjutnya pembersihan data, pembentukan data baru, proses data mining, dan yang terakhir adalah evaluasi.

- a. Identifikasi dan Pengumpulan Data Pada tahap ini dilakukan identifikasi terhadap penelitian yang akan dilakukan dan melakukan pengumpulan data yang sesuai dengan penelitian. Pada pengumpulan data, data yang digunakan adalah data Penerimaan Mahasiswa Baru tahun 2016.
- b. Pencarian Literatur Tahap ini adalah mencari literatur dari buku-buku maupun jurnal penelitian terdahulu tentang penerapan Algoritma k-Nearest Neighbor, dan algoritma untuk pencarian k-Optimal yaitu k-Fold Cross Validation.
- c. Persiapan dan Pemilihan Data Melakukan persiapan terhadap data yang telah didapat seperti melihat struktur tabel yang ada pada database. Pemilihan data dilakukan karena tidak semua tabel serta data yang ada dalam database berhubungan dengan penelitian yang dilakukan, sehingga hanya data yang berkaitan dengan penelitian yang akan digunakan.
- d. Pembersihan Data Tahap ini dilakukan untuk memastikan bahwa tidak ada data yang terduplikasi, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Data yang telah bersih dari kesalahan dapat

mempermudah penelitian dan mencegah adanya kesalahan pada penelitian.

- e. Pembentukan Data Baru Pembentukan data baru ini agar data yang didapat dan telah bersih dari kesalahan bisa dibentuk menjadi sebuah tabel baru yang sesuai dengan algoritma kNN untuk menemukan pola penerimaan mahasiswa baru serta mencari k-Optimal pada algoritma kNN.
- f. Proses Data Mining Tahap yang sangat penting dalam penelitian. Pada tahap ini ada beberapa tahap dilakukan yaitu: 1) Proses pencarian k-Optimal pada algoritma kNN menggunakan metode kFold Cross Validation. 2) Setelah mendapatkan nilai k-Optimal, nilai tersebut digunakan untuk di uji akurasi menggunakan data real sebanyak 166 buah data (30%) dari database sehingga akan diketahui berapa banyak data yang memiliki banyak ketepatan prediksi.
- g. Tahap terakhir adalah melakukan prediksi kelulusan tepat waktu mahasiswa dengan menggunakan algoritma kNN dengan nilai k hasil dari k-Folds Cross Validation dengan variabel input yaitu mahasiswa yang melakukan her-registrasi.
- h. Evaluasi Tahap ini adalah tahap dimana pola informasi yang dihasilkan dari proses data mining ditampilkan dalam bentuk yang dapat dipahami oleh pihak yang berkepentingan.

II Tinjauan Pustaka

2.1 Data Mining

Data mining atau penggalian data merupakan proses yang memperkerjakan satu atau lebih teknik (pola) pembelajaran komputer untuk mengekstraksi serta menganalisis suatu pengetahuan dengan cara otomatis. Definisi lain ialah pembelajaran berbasis induksi (induction-based learning) merupakan proses pembentukan definisi konsep umum yang dilakukan dengan cara mengobservasi contoh spesifik dari konsep-konsep yang akan dipelajari. Knowledge Discovery in Database (KDD) merupakan sebuah penerapan metode saintifik pada data mining. Dalam konteks ini data mining adalah satu langkah dari proses KDD. Data mining mempunyai beberapa nama alternatif, walaupun definisi eksaknya berbeda, KDD, analisis pola, arkeologi data, intelegensia bisnis, serta pemanenan informasi. Data mining diperlukan pada saat data tersedia terlalu banyak

(contohnya data yang didapat dari sistem basis data perusahaan, e-commerce, data bioinformatika, serta data saham), namun tidak tahu pola apa yang bisa didapatkan.

Perkembangan yang sangat pesat pada bidang pengumpulan data serta teknologi penyimpanan di berbagai bidang, yang dapat menghasilkan basis data yang terlalu besar. Tetapi, data yang sudah dikumpulkan jarang dilihat lagi, sebab terlalu panjang, membosankan, serta tidak menarik. Sering, keputusan yang katanya berdasarkan data- dibuat tidak lagi berdasarkan data, namun dari intuisi para pembuat keputusan, sehingga, lahirilah cabang ilmu penggalian data tersebut.

Analisis data tanpa memakai otomatisasi dari penggalian data merupakan hal yang tidak memungkinkan lagi, jika 1) data terlalu banyak 2) dimensionalitas data terlalu besar, 3) data terlampaui kompleks untuk dianalisis manual (contohnya: data time series, data spatiotemporal, data multimedia, data streams).

2.2 Proses Pencarian Pola

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantaraan knowledge base. Tahapan-tahapan tersebut, diantaranya:

a. Pembersihan data

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Garbage in garbage out (hanya sampah yang akan dihasilkan bila yang dimasukkan juga sampah) merupakan istilah yang sering dipakai untuk menggambarkan tahap ini. Pembersihan data juga akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

b. Integrasi data

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dsb. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya.

c. Transformasi data

Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut binning. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.

d. Aplikasi teknik data mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai. Pembahasan lebih jauh mengenai teknik-teknik yang ada di seksi berikutnya. Perlu diperhatikan bahwa ada kalanya teknik-teknik data mining umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data mining di bidang tertentu atau untuk data tertentu. Sebagai contoh akhir-akhir ini dikembangkan berbagai teknik data mining baru untuk penerapan di bidang bioinformatika seperti analisa hasil microarray untuk mengidentifikasi DNA dan fungsi-fungsinya.

e. Evaluasi pola yang ditemukan

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi

dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti : menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

Presentasi pola yang ditemukan untuk menghasilkan aksi

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

2.3. Fungsi umum data mining

Fungsi dalam data mining bisa diartikan sebagai sub kegiatan yang ada dalam data mining dalam rangka menemukan, menggali, atau menambang pengetahuan. Secara global fungsi utama data mining adalah sebagai berikut .

1. Klasifikasi

Proses untuk menyatakan suatu objek ke dalam satu kategori yang sudah didefinisikan sebelumnya. Proses pembelajaran fungsi target (model klasifikasi) yang memetakan setiap sekumpulan atribut x (input) ke dalam satu kelas yang didefinisikan sebelumnya.

Input: sekumpulan record (training set).

Setiap record terdiri atas sekumpulan atribut, salah satu atribut adalah kelas. Mencari model untuk atribut kelas sebagai fungsi dari nilai-nilai untuk atribut yang lain.

Tujuannya adalah record-record yang sebelumnya tidak terlihat dinyatakan kelasnya seakurat mungkin. Ada banyak sekali algoritma yang terdapat pada teknik klasifikasi , beberapa diantaranya seperti :

- Holte

- Prism
- Naïve Bayes
- ID3
- K-Nearest Neighbor

2. Pengelompokan / Clustering

Clustering adalah metode penganalisaan data yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain.

Perbedaan clustering dengan classification

Dataset yang digunakan pada clustering tidak menampilkan class / target attribute, sedangkan dataset yang digunakan pada classification mutlak harus menampilkan class / target attribute.

Pengetahuan yang dihasilkan oleh metode clustering berupa cluster hasil pengelompokan, sedangkan pengetahuan yang dihasilkan oleh metode classification berupa selain cluster (bisa Decision Tree, Ruleset, Weight pada BackPropagation, dan lain-lain).

Clustering dipakai ketika tidak diketahuinya bagaimana data harus dikelompokkan. Jumlah kelompok diasumsikan sendiri tanpa ditentukan terlebih dahulu.

Keluaran pendekatan ini adalah data yang sudah dikelompokkan. Sedangkan classification, terdapat informasi mengenai bagaimana data tersebut dikelompokkan.

Kemudian dilakukan training pada sistem dengan data yang sudah diberikan label (ke dalam kelompok manakah data tersebut dikelompokkan), selanjutnya sistem akan mengklasifikasikan data-data yang baru ke dalam kelompok yang ada. Tidak akan ada penambahan kelompok.

Algoritma yang sering di pakai untuk pengelompokan data adalah Algoritma K-Means.

3. Asosiasi

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item.

Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu.

Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah market basket analysis.

Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya.

Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (frequent pattern mining) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien.

Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, support (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Aturan asosiatif biasanya dinyatakan dalam bentuk :

{roti, mentega} _ {susu} (support = 40%, confidence = 50%)

Yang artinya : “50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu.”

Dapat juga diartikan: “Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini.” Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (minimum support) dan syarat minimum untuk confidence (minimum confidence).

Algoritma yang biasa digunakan untuk analisis asosiasi adalah Algoritma A Priori.

Penerapan data mining sebagai cabang dari bidang ilmu komputer sudah cukup banyak digunakan dalam bidang industry bisnis dan lainnya. Sehingga bisa sangat membantu jika di gunakan dengan tepat.

Demikian ulasan tentang pengertian data mining dan konsepnya dari saya, semoga bisa membantu dan menambah pengetahuan anda tentang data mining.

2.4. K-Nearest Neighbor

Algoritma k-nearest neighbor (k-NN atau KNN) 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. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidean.

Pada fase pembelajaran, algoritme ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data test (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah k buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut.

Nilai k yang terbaik untuk algoritme ini tergantung pada data; secara umumnya, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan cross-validation. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, k = 1) disebut algoritme nearest neighbor.

Ketepatan algoritme k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritme ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi menjadi lebih baik.

Terdapat beberapa jenis algoritme pencarian tetangga terdekat, diantaranya:

- Linear scan
- Pohon kd
- Pohon Balltree
- Pohon metrik
- Locally-sensitive hashing (LSH)

Algoritme k-NN ini memiliki konsistensi yang kuat. Ketika jumlah data mendekati tak hingga, algoritme ini menjamin error rate yang tidak lebih dari dua kali Bayes error rate (error rate minimum untuk distribusi data tertentu).

IV. HASIL DAN PEMBAHASAN

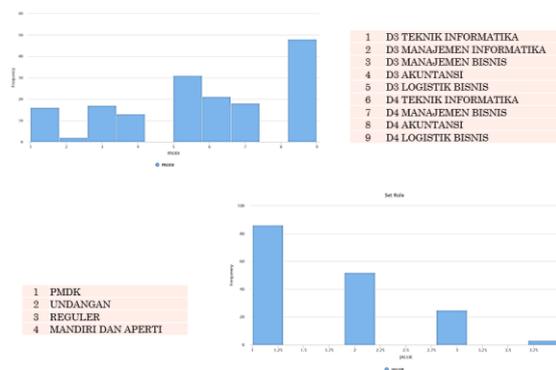
Pembahasan hasil perhitungan pada setiap kluster dapat di informasikan sebagai berikut: Pada kluster pertama mahasiswa didominasi oleh Program Studi D4 Logistik Bisnis, yang kebanyakan berasal dari SMA Negeri yang berasal dari cluster kota 1-2 dan 9-10 meliputi Kota, Kabupaten Bandung, dan Bandung Barat, Garut, Tasikmalaya sebagaimana yang ditunjukkan pada gambar 1.



Gambar 1. Profil Penerimaan Mahasiswa Baru.

Selain profil sekolah, prodi dan kota yang dihasilkan dalam analisis data juga dihasilkan profil Jalur pendaftaran paling banyak melalui Jalur

PMDK dan Undangan sebagaimana yang tersajikan pada gambar 2.



Gambar 2. Profil Jalur Pendaftaran dan Pilihan Prodi

Secara keseluruhan pola yang dihasilkan dari proses data mining Peneruman mahasiswa Baru Politeknik pos Indonesia pada tahun 2016 memberikan informasi baru dalam mendukung strategi promosi penerimaan mahasiswa baru. Secara umum data mining dengan menggunakan algoritma K-Nearest Neighbor mendapatkan tingkat kepercayaan sebesar 67% – 77%. Hal ini ditunjukkan pada hasil dibawah ini.

Row No.	STATUS	prediction(STATUS)	JALUR	PRODI	SEKOLAH	KOTA	squared_error
1	1	1	2	9	1	1	squared_error: 0.239 +/- 0.332
2	1	1	1	4	2	3	
3	1	1	2	2	1	1	
4	1	1	3	4	1	1	
5	1	1	2	9	1	7	
6	0	0	1	4	1	7	
7	1	1	1	7	9	4	

Gambar 3. Hasil testing data PMB 2016

V KESIMPULAN

Beberapa simpulan yang dapat disajikan adalah:

1. Pola pembelajaran mesin dari data PMB tahun 2016 memberikan hasil yang cukup baik dengan tingkat kepercayaan 67% – 77%
2. Data Penerimaan mahasiswa baru (PMB) setiap tahun dapat dijadikan ajuan sebagai informasi yang digunakan untuk mendapatkan model penerimaan mahasiswa baru di Politeknik Pos Indonesia.
3. Tim PMB dapat mengetahui Profil mahasiswa yang akan melakukan Her Registrasi di kampus Politeknik Pos Indonesia
4. Strategi promosi dan penerimaan mahasiswa baru dapat dilakukan secara ilmiah dengan

mengacu pada model dan hasil analisis data PMB.

DAFTAR PUSTAKA

- [1] Larose, Daniel T, "Discovery Knowledge in Data: An Introduction To Data Mining", Jhon Willey & Sons, Inc, 2005.
- [2] Turban, E.,dkk, "Decicion Support System and Inteligent System", Yogyakarta: Andi Offset, 2005.
- [3] Darmawan, Astrid. 2012. Pembuatan Aplikasi Data mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighborhood. Fakultas Teknik Dan Ilmu Komputer Universitas Komputer Indonesia : Bandung
- [4] Pandie, Emerensye S. Y. 2012. Implementasi Algoritma Data mining KNearest Neighbour (KNN) Dalam Pengambilan Keputusan Pengajuan Kredit. Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana : Kupang.

Ketentuan Penulisan Jurnal Ilmiah Manajemen Informatika

Umum

Jurnal Ilmiah Manajemen Informatika menerima karya tulis:

1. Dalam bentuk hasil penelitian, tinjauan pustaka dan laporan kasus dalam bidang ilmu yang berhubungan dengan teknologi informasi.
2. Belum pernah dipublikasikan dalam majalah / jurnal ilmiah manapun. Bila pernah dipresentasikan, sertakan keterangan acara, tempat dan tanggalnya.
3. Ditulis dalam bahasa Indonesia.

Sistematika yang diterapkan untuk tiap kategori karya-karya tersebut adalah:

1. Hasil penelitian
Hasil penelitian terdiri atas judul, penulis, abstrak berbahasa Indonesia (terdiri dari 150 – 200 kata), disertai kata kuncinya. Pendahuluan, metode, pembahasan, kesimpulan dan saran, serta daftar pustaka (merujuk sekurang-kurangnya tiga pustaka terkini)
2. Tinjauan pustaka
Naskah hasil studi literatur terdiri atas judul dan penulis. Pendahuluan (disertai pokok-pokok ide kemajuan pengetahuan terakhir sehubungan dengan masalah yang digali). Permasalahan mencakup rangkuman sistematika dari berbagai narasumber. Pembahasan menurut ulasan dan sintesis ide. Kesimpulan dan saran disajikan sebelum daftar pustaka. Tinjauan pustaka merujuk pada sekurang-kurangnya tiga sumber pustaka terbaru.
3. Laporan kasus
Naskah laporan kasus terdiri atas judul, abstrak berbahasa Indonesia (terdiri dari 50-100 kata) disertai kata kuncinya, pendahuluan (disertai karakteristik lokasi, gambaran umum budaya yang relevan, dll), masalah pembahasan dan resume atau kesimpulan.

Format

Naskah hendaknya ditulis singkat, padat, konsisten, dan lugas. Jurnal tidak akan memuat naskah dengan jumlah halaman lebih dari 20 (dua puluh). Naskah ditulis dalam spasi tunggal pada satu sisi kertas ukuran A4 (210 x 297 mm), dengan margin atas dan bawah 2,5 cm serta margin kiri 3 cm dan margin kanan 2,5 cm. Huruf yang digunakan adalah *Time New Roman* 10 pt, dibuat dalam 2 (dua) kolom. Naskah dapat ditulis dengan menggunakan bahasa Indonesia atau bahasa Inggris yang baik dan benar.

Judul dan Abstrak

Judul hendaknya dibuat singkat, padat, dan mencerminkan isi naskah keseluruhan. Judul ditulis ditengah-tengah, huruf yang digunakan adalah *Time New Roman* 12 pt. Dibawah judul dituliskan nama (para) penulis. Dibawah nama dituliskan afiliasi dari (para) penulis, dan diikuti dengan alamat e-mail (para) penulis.

Abstrak ditulis dalam bahasa Indonesia atau Inggris dan dalam bentuk satu kolom. Sedapat mungkin abstrak tidak berisikan rumus dan referensi. Abstrak harus ringkas, tujuan, lingkup, hasil utama, dan kesimpulan penelitian. Panjang abstrak maksimum adalah 200 kata. Abstrak dilengkapi dengan kata kunci kurang lebih 4-6 buah.

Rumus, Gambar, dan Tabel

Setiap rumus diberi nomor pemunculan di sisi kanan dengan menggunakan angka Arab di dalam kurung. Sedangkan setiap tabel dan gambar diberi nomor menggunakan angka Arab disertai dengan keterangan (judul). Nomor dan keterangan tabel diletakkan di atas tabel sedangkan nomor dan keterangan gambar diletakkan di bawah gambar dengan posisi di tengah (*center*).

Daftar Pustaka

Setiap rujukan disertai dengan keterangan yang mengacu pada daftar pustaka. Keterangan ini berupa nama penulis dan tahun publikasi. Contoh: (Wheelwright dan Clark, 1992), (Whitney, 1998), (Simatupang et al., 2004).

Semua referensi yang digunakan ditulis pada daftar pustaka dengan contoh format sebagai berikut:

Wheelwright, S.C dan Clark, K.B (1992). *Revolutioning Product Development*. The Free Press, New York.

Whitney, D.E. (1998), "Manufacturing by design", *Harvard Business Review*. Vol. 66 No. 3. pp. 83-91.

Simatupang, T.M., Sandroto, I.V. dan Lubis, S.B.H. (2004), "A Coordination Analysis of the Creative Design Process", *Business Process Management Journal*. Vol. 10 No.4 pp.430-444.