

INTEGRASI SAMPLE BOOTSTRAPPING PADA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI HERREGISTRASI CALON MAHASISWA BARU

Eny Jumiati⁽¹⁾ Muhammad Rizkam Kamal⁽²⁾

Program Studi Teknik Informatika STMIK Widya Pratama

Jl. Patriot 25 Pekalongan Telp (0285) 427816

⁽¹⁾Email: enijumiati@stmik-wp.ac.id

⁽²⁾Email: rikzamrx@gmail.com

ABSTRAK

Herregistrasi calon mahasiswa baru adalah proses pendaftaran ulang setiap calon mahasiswa baru untuk dapat mengikuti kegiatan perkuliahan, memperoleh hak-hak akademik, dan menggunakan fasilitas bagi mahasiswa sebagaimana ketentuan yang berlaku pada semester yang berjalan. Dari data pendaftaran mahasiswa baru yang dibuka setiap tahunnya didapatkan banyak calon mahasiswa baru yang mendaftar, tetapi dari data herregistrasi tidak semua calon mahasiswa baru yang sudah mendaftar melakukan herregistrasi. Data mining adalah proses menemukan pola yang menarik dari sejumlah data besar. Sebagai proses penemuan pengetahuan, biasanya melibatkan data cleaning, integrasi data, seleksi data, transformasi data, penemuan pola, dan evaluasi pola. Untuk menangani masalah pengklasifikasian data pada data herregistrasi calon mahasiswa baru dapat dilakukan dengan mengintegrasikan sample bootstrapping ke dalam k-NN (k-NNSB) untuk meningkatkan akurasi hasil klasifikasi. Didapatkan hasil akurasi sebesar 88.74% pada nilai variabel k-2.

Kata kunci: Herregistrasi, K-NN, Sample Bootstrapping

1. Pendahuluan

1.1. Latar Belakang

Mahasiswa adalah sebutan bagi orang yang sedang menempuh pendidikan tinggi di sebuah perguruan tinggi yang terdiri atas sekolah tinggi, akademi, dan yang paling umum adalah universitas. Herregistrasi calon mahasiswa baru adalah proses pendaftaran ulang setiap calon mahasiswa baru untuk dapat mengikuti kegiatan perkuliahan, memperoleh hak-hak akademik, dan menggunakan fasilitas bagi mahasiswa sebagaimana ketentuan yang berlaku pada semester yang berjalan. Dari data pendaftaran mahasiswa baru yang dibuka setiap tahunnya didapatkan banyak calon mahasiswa baru yang mendaftar, tetapi dari data herregistrasi tidak semua calon mahasiswa baru yang sudah mendaftar melakukan herregistrasi.

Data mining adalah proses menemukan pola yang menarik dari sejumlah data besar. Sebagai proses penemuan pengetahuan, biasanya melibatkan data cleaning, integrasi data, seleksi data, transformasi data, penemuan pola, dan evaluasi pola (Han 2012). Pada perkembangannya data mining dipergunakan dalam bidang ekonomi, statistika, peramalan, dan insinyur komunikasi yang telah lama bekerja sama dengan gagasan bahwa pola dalam data dapat dicari secara otomatis, diidentifikasi, divalidasi, dan digunakan untuk prediksi (Witten, Frank & M. a. Hall 2011). Di

dalam data mining terdapat 10 algoritma teratas yang biasa dipakai oleh para peneliti dalam bidang data mining, 6 algoritma diantaranya adalah k-Nearest Neighbor (k-NN), C4.5, Support Vector Machines (SVM), AdaBoost, Naïve Bayes, dan CART (Witten, Frank & M. A. Hall 2011), yang mana ini adalah termasuk ke dalam algoritma klasifikasi.

K-NN adalah teknik sederhana namun sangat kuat untuk mengklasifikasikan berbagai data (Heathcote 2013). K-NN dimana tetangga terdekat dihitung atas dasar nilai k , yang menentukan berapa banyak tetangga terdekat yang harus dipertimbangkan untuk menentukan kelas dari data titik sampel (Bhatia & Vandana 2010). K-NN didasarkan pada fungsi jarak yang mengukur perbedaan atau kesamaan antara dua contoh. K-NN adalah contoh pada *lazy learning* (Jiang et al. 2007). *Lazy learning* hanya menyimpan data pelatihan pada saat pelatihan dan penundaan belajar sampai waktu klasifikasi. Sebaliknya *eager learning* menghasilkan model eksplisit pada saat pelatihan.

Beberapa peneliti telah melakukan penelitian mengenai klasifikasi tentang herregistrasi calon mahasiswa baru, seperti yang dilakukan oleh (Al Karomi 2015) yang menggunakan k-NN untuk mengklasifikasi data herregistrasi

mahasiswa, dimana dilakukan optimasi pemilihan nilai k untuk mendapatkan nilai yang paling optimal.

Untuk menangani masalah pengklasifikasian data pada data herregistrasi calon mahasiswa baru dapat dilakukan dengan mengintegrasikan sample bootstrapping ke dalam k -NN (k -NNSB) untuk meningkatkan akurasi hasil klasifikasi.

1.2. Landasan Teori

1.2.1. Data Mining

Data mining merupakan sebuah proses ekstraksi untuk mendapatkan suatu informasi yang sebelumnya tidak diketahui dari sebuah data (Witten, Frank & M. a. Hall 2011). Data Mining atau sering juga disebut Knowledge Discovery in Database (KDD) adalah sebuah bidang ilmu yang banyak membahas tentang pola sebuah data. Serangkaian proses guna mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan data mining. Beberapa peran utama data mining adalah: Estimation, Prediction, Classification, Clustering dan Association. Dari semua peranan data mining tersebut terbagi menjadi 2 berdasarkan metode pembelajarannya yaitu Supervised Learning, Unsupervised Learning. Perbedaan dari kedua metode pembelajaran pada algoritma data mining tersebut adalah jika dalam supervised learning harus memiliki data sampel atau sering disebut juga dengan data training. Sedangkan dalam unsupervised learning tidak membutuhkan data training. Salah satu contoh peran data mining dengan metode supervised learning adalah klasifikasi.

1.2.2. K-Nearest Neighbor (k -NN)

k -NN merupakan algoritma *supervised learning*, dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k -NN. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi. k -NN merupakan salah satu metode pengklasifikasian data berdasarkan similaritas dengan label data, Algoritma k -NN merupakan sebuah metode untuk melakukan klasifikasi terhadap objek baru berdasarkan k tetangga terdekatnya. Menurut M. Han, J., & Kamber (Han 2012), k -NN adalah algoritma pembelajaran berbasis instan yang

menggunakan jarak terdekat dalam menentukan kategori vektor baru dalam set data training.

Model k -NN memiliki masalah untuk menemukan tetangga terdekat k pada titik query dari dataset yang digunakan. Penggunaan metode ini banyak digunakan untuk menangani masalah dalam bidang ilmiah dan rakayasa perangkat lunak seperti pengenalan pola, pengenalan objek, pengelompokan data, fungsi approximate, kuantisasi vektor, klasifikasi pola.

Untuk menentukan jumlah data atau tetangga terdekat ditentukan oleh user yang dinyatakan dengan k .

$$\text{dist}(x_1, x_2) = \sqrt{\sum_{i=1}^n (X_{1i} - X_{2i})^2}$$

dimana $X_1 = (X_{11}, X_{12}, \dots, X_{1n})$ dan $X_2 = (X_{21}, X_{22}, \dots, X_{2n})$. Dengan kata lain, untuk setiap atribut numerik, kita mengambil perbedaan antara yang sesuai nilai-nilai atribut yang dalam vektor X_1 dan X_2 dari matriks dengan ukuran dimensi. Akar kuadrat diambil dari akumulasi jumlah total jarak. Biasanya, kita menormalkan nilai masing-masing atribut sebelum digunakan. Prinsip kerja k -Nearest Neighbor (k -NN) adalah mencari jarak terdekat antara data yang dievaluasi dengan k tetangga terdekatnya dalam data pelatihan. Persamaan Penghitungan untuk mencari euclidean dengan d adalah jarak dan p adalah dimensi data dengan:

$$d_i = \sqrt{\sum_{i=1}^p (X_{1i} - X_{2i})^2}$$

dimana:

x_1 : sample data uji

x_2 : data uji

d : jarak

p : dimensi data

1.2.3. Sample Bootstrapping

Untuk melakukan estimasi parameter dapat menggunakan metode resampling seperti bootstrapping ataupun jackknife sebagai prosedur alternatif untuk pendekatan non-parametrik yang memerlukan asumsi dan kesulitan untuk menilai suatu parameter. Dengan prosedur bootstrapping bersifat

estimator, seperti variansnya dapat diperkirakan melalui pengambilan sampel secara berulang dengan penggantian distribusi silang seperti distribusi empiris dari sampel observasi. Penggunaan metode resampling dapat digunakan untuk aplikasi klasifikasi dengan metode pembobotan terhadap pencarian tetangga terdekat.

Metode resampling, seperti jackknife dan bootstrapping juga telah digunakan untuk menghitung kinerja klasifikasi. Bootstrapping bekerja dengan menghasilkan dataset baru dari dataset asli dengan resampling dengan atau tanpa penggantian, dan kemudian masing-masing dataset baru digunakan untuk menghitung parameter bootstrapping dari proses klasifikasi. Penentuan parameter akhir pada kinerja klasifikasi untuk dataset akan dievaluasi agar diperoleh rata-rata parameter bootstrapping.

Bootstrapping merupakan suatu metode yang dapat bekerja tanpa membutuhkan asumsi distribusi karena sampel asli digunakan sebagai populasi. Metode bootstrapping adalah teknik resampling nonparametrik yang bertujuan untuk menentukan estimasi standard error dan interval konfidensi dari parameter populasi seperti mean, rasio, median, proporsi, koefisien korelasi atau koefisien regresi tanpa menggunakan asumsi distribusi.

Bootstrapping merupakan suatu metode yang berbasis komputer yang sangat potensial untuk dipergunakan pada masalah keakuratan. Bootstrapping dapat digunakan untuk mengatasi permasalahan dalam statistika baik masalah data yang sedikit, data yang menyimpang dari asumsinya maupun data yang tidak memiliki asumsi dalam distribusinya.

Menurut Efron dan Tibshirani (B. Efron 2011), prosedur resampling bootstrap dapat dituliskan sebagai sebagai berikut:

1. Mengkonstruksi distribusi empiris n dari F dari suatu sampel dengan memberikan propabilitas $1/n$ pada setiap X_i dimana $i = 1, 2, \dots, n$.

2. Mengambil sampel bootstrap berukuran n secara random dengan pengembalian dari distribusi tahap 1, distribusi empiris n dan F , disebut sebagai sample bootstrap X^*1 .
3. Menghitung dari tahap 1, statistik $\hat{\theta}$ yang diinginkan dari sample bootstrap, disebut sebagai $\hat{\theta}^*1$.
4. Mengulangi langkah 2 dan 3 hingga B kali, diperoleh $\theta^*1, \theta^*2, \dots, \theta^*B$.
5. Mengkonstruksi suatu distribusi probabilitas dari \hat{F} dengan memberikan probabilitas $1/B$ pada setiap $\theta^*1, \theta^*2, \dots, \theta^*B$. Distribusi tersebut merupakan estimator bootstrap untuk distribusi sampling θ dan \hat{F} dinotasikan dengan \hat{F}^* .
6. Pendekatan estimasi bootstrap untuk mean dari 1, distribusi yaitu:

$$\hat{\theta} = \sum_{b=1}^B \hat{\theta}_b^* \frac{1}{B}$$

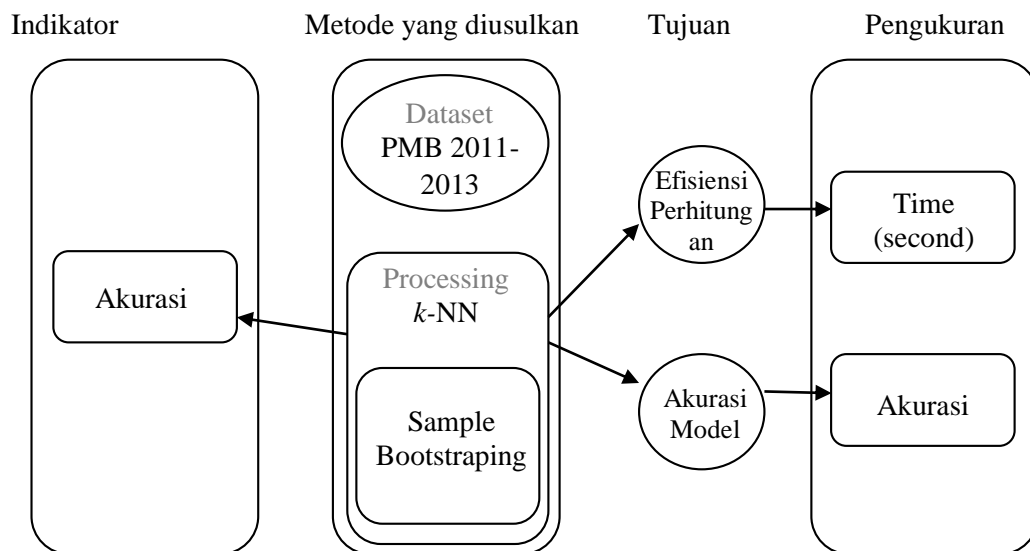
1.2.4. Cross Validation

Dalam *data mining cross validation* adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Dengan menggunakan cross validation akan dilakukan percobaan sebanyak k . Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan.

Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi seperti pada Gambar 1. Ilustrasi 10-Cross Fold Validation. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau *10-cross fold validation*, setiap percobaan akan menggunakan satu data testing dan $k-1$ bagian akan menjadi data *training*, kemudian data testing itu akan ditukar dengan satu buah data training sehingga untuk setiap percobaan akan didapatkan data *testing* yang berbeda-beda.

Split1	Split2	Split3	Split4	Split5	Split6	Split7	Split8	Split9	Split10
Training									Test
Training								Test	
Training							Test		
Training						Test			
Training					Test				
Training				Test	Training				
Training			Test	Training					
Training		Test	Training						
Test									

Gambar 1. Ilustrasi 10-Cross Fold Validation



Gambar 2. Kerangka Pemikiran Penelitian

1.2.5. Kerangka Pemikiran

Penelitian ini secara garis besar mempunyai kerangka pemikiran mulai dari indikator, metode yang diusulkan, tujuan dan pengukuran. Kemudian untuk dataset yang digunakan pada penelitian ini adalah dataset Penerimaan Mahasiswa Baru (PMB) 2011 sampai 2013 STMIK Widya Pratama Pekalongan.

Pada Gambar 2. menunjukkan kerangka pemikiran penelitian ini dengan

mengintegrasikan *Sample Bootstrapping* ke dalam *k-NN* untuk meningkatkan akurasi dan waktu komputasi dalam pengklasifikasi dataset PMB 2011-2013 (herregistrasi calon mahasiswa baru). Metode *Sample Bootstrapping* digunakan untuk mengurangi jumlah data training yang akan diproses dengan memberikan nilai parameter pada sample ratio (n). Untuk pengujian akurasi hasil klasifikasi dilakukan menggunakan metode confusion matrix dan uji efisiensi (lamanya waktu proses klasifikasi) dinyatakan dalam waktu (detik).

2. Metode Penelitian

2.1. Tahapan Penelitian

Metode penelitian yang digunakan dalam penelitian ini seperti pada Gambar 3.. Tahapan penelitian yang di dalamnya berisi tentang tahapan yaitu:

1. Pengumpulan data

Pada penelitian ini diawali dengan mengumpulkan data. Data didapatkan dari PMB (Penerimaan Mahasiswa Baru) 2011-2013 STMIK Widya Pratama Pekalongan.

2. Metode

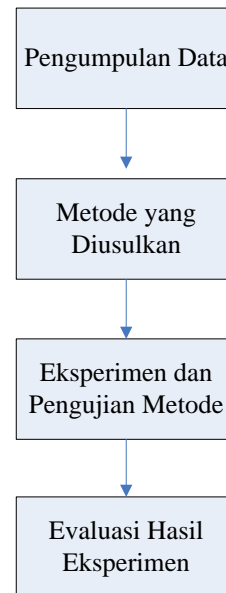
Metode yang digunakan dalam penelitian ini berupa metode dalam teknik klasifikasi data mining yang dioptimisasi untuk mengurangi jumlah data training dan mengurangi atribut.

3. Eksperimen dan Pengujian Model

Untuk melakukan pengujian model, dilakukan dengan menggunakan Rapid Miner, dari algoritma yang sudah ditentukan maka selanjutnya dataset yang sudah ada akan diolah sehingga menghasilkan model yang akan diinginkan.

4. Evaluasi dan Validasi Hasil

Setelah melakukan eksperimen terhadap semua dataset dengan model yang diusulkan maka akan menghasilkan nilai-nilai akurasi dan waktu komputasi dari model yang digunakan kemudian hasil tersebut dianalisa dan dievaluasi. Dari hasil evaluasi selanjutnya dapat ditarik kesimpulan dari penelitian dan eksperimen ini.



Gambar 3. Tahapan Penelitian

2.2. Pengumpulan Data

Pada penelitian ini menggunakan dataset PMB 2011-2013 STMIK Widya Pratama Pekalongan seperti pada Tabel 1. Adapun data yang digunakan adalah berupa 2389 baris dan 32 kolom.

Tabel 1. Dataset PMB 2011-2013

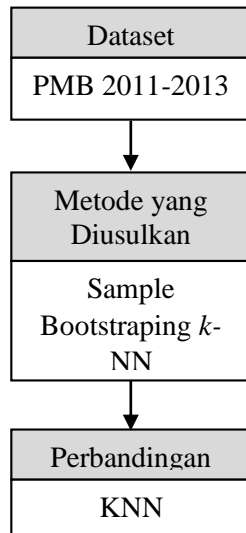
Nama	Tipe	Record	Kolom
PMB 2011- 2013	Klasifikasi	2389	32

2.3. Metode yang Diusulkan

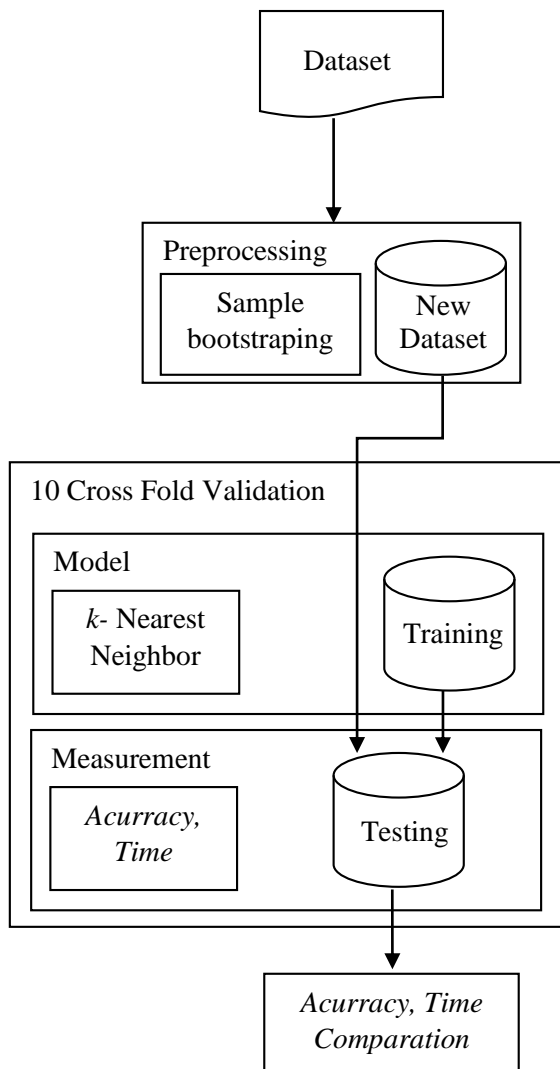
Menggambarkan metode yang digunakan serta menerangkan bagaimana metode yang diusulkan untuk menyelesaikan masalah penelitian, yang dapat dilihat pada Gambar 4., dimana *Sample Bootstrapping* digunakan untuk mengurangi jumlah data training yang akan proses.

Pada Gambar 5. menunjukan proses *Sample Bootstrapping* dan *k-NN*. Pertama dari dataset yang dilakukan proses *preprocessing* dengan menggunakan *Sample Bootstrapping* agar mengurangi jumlah data yang nantinya akan digunakan untuk proses training. Pada proses

validasi menggunakan *cross validation*, setelah itu hasil dari perhitungan diukur tingkat akurasi dan waktu komputasinya.



Gambar 4. Metode yang Diusulkan



Gambar 5. Integrasi Sample Bootstrapping ke dalam *k*-NN

2.4. Eksperimen dan Pengujian Metode

Dalam eksperimen ini mempunyai tahapan sebagai berikut:

1. Menyiapkan dataset yang digunakan
2. Melakukan pengujian menggunakan *k*-NN standar dengan dataset PMB 2011-2013 STMIK Widya Pratama Pekalongan.
3. Melakukan pengujian *k*-NNBS dengan menggunakan dataset PMB 2011-2013 STMIK Widya Pratama Pekalongan.
4. Melakukan komparasi hasil uji dari *k*-NN standar dengan hasil uji yang dihasilkan pada *k*-NN dengan Sample Bootstrapping.

Spesifikasi komputer yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2.

Tabel 1. Spesifikasi Komputer yang Digunakan

Processor	Intel (R) Core (TM) 2 Quad CPU Q8300 @2.50GHz 2.49 GHz
Memori	DDR3 4 GB
Hardisk	500 GB
Sistem Operasi	Windows 7 Professional 64- bit
Aplikasi	RapidMiner 5.3.015.

2.5. Evaluasi Hasil Eksperimen

Adapun model yang dihasilkan akan diuji dengan menggunakan *cross validation* untuk mengetahui tingkat akurasi, *cross validation* digunakan untuk menghindari overlapping pada data testing. Tahapan *cross validation* terdiri dari proses pembagian data menjadi *k* subset yang berukuran sama kemudian menggunakan setiap subset untuk data testing dan sisanya untuk data training. Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan. Pengujian juga diukur dengan menggunakan ROC (Receiver Operating Characteristic) Curve. ROC Curve akan

menggambarkan kelas positif dalam bentuk kurva. Pengujian dilakukan dengan menghitung nilai AUC (Area Under Curve), semakin tinggi nilai AUC dalam ROC Curve, maka semakin baik pula model klasifikasi yang terbentuk. Sedangkan untuk pengujian akurasi hasil klasifikasi dilakukan menggunakan metode confusion matrix dan uji efisiensi (lamanya waktu proses klasikasi) dinyatakan dalam waktu (detik) dengan formula:

$$\text{Akurasi} = \frac{\text{Jumlah data yang benar}}{\text{Jumlah data}} \times 100\%$$

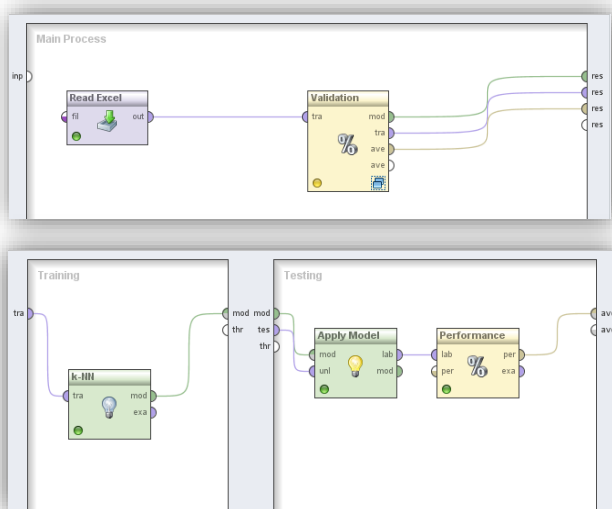
3. HASIL DAN PEMBAHASAN

3.1. Hasil

Pada penelitian ini dilakukan penelitian terhadap herregistrasi calon mahasiswa baru STMIK Widya Pratama Pekalongan dengan mengintegrasikan sample bootstrapping pada *k*-nearest neighbor (*k*-NN). Untuk melakukan penelitian ini dilakukan beberapa hal dengan menggunakan bantuan aplikasi rapidminer 5.3.015.

3.1.1. Mendesain K-NN

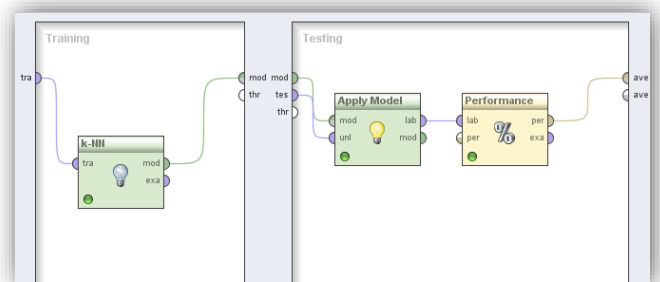
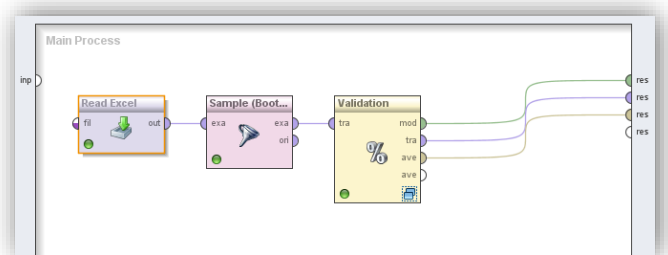
Mendesain *k*-NN standar yang digunakan untuk mengolah dataset PMB 2011-2013 STMIK Widya Pratama Pekalongan. Yang nantinya akan difungsikan untuk mengklasifikasin data tersebut. Dimana desainnya seperti pada Gambar 6..



Gambar 6. Desain Metode *k*-NN

3.1.2. Mendesain K-NN Bootstrapping

Pada penelitian ini selain menggunakan *k*-NN yang standar, kita juga mencoba mendesain *k*-NN *Bootstrapping*. Seperti pada Gambar 7. yang menjelaskan desain yang digunakan, dimana ditambahkan bootsrapping yang digunakan sebagai metode resampling data yang akan diolah yang kemudian akan kita masukan ke dalam metode *k*-NN untuk proses klasifikasi.



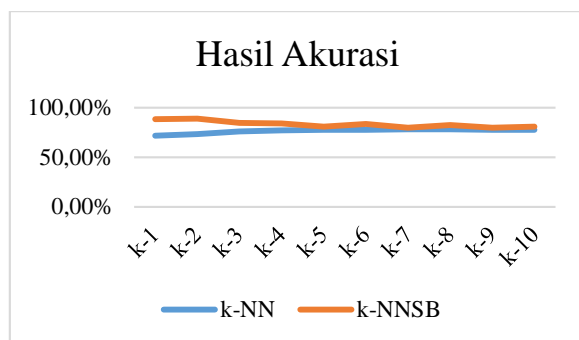
Gambar 7. Desain Metode *k*-NN

3.1.3. Hasil Akurasi

Setelah dilakukan beberapa pengujian dihasilkan beberapa hasil, diantaranya adalah tentang akurasi pada k -NN dan juga k -NNSB yang digunakan untuk mengkalsifikasi data PMB 2011-2013. Dimana pengujian ini dilakukan dengan menggunakan beberapa nilai pada variabel k . Dari $k-1$ sampai dengan $k-10$ seperti pada Tabel 3.

Tabel 3. Hasil Akurasi

Parameter k	Akurasi		Selisih
	k -NN	k -NN Sample Bootstraping	
$k-1$	71.79%	88.32%	16.53%
$k-2$	73.04%	88.74%	15.7%
$k-3$	75.72%	84.47%	8.75%
$k-4$	76.77%	83.93%	7.16%
$k-5$	77.61%	80.49%	2.88%
$k-6$	77.27%	83.38%	6.11%
$k-7$	78.15%	79.87%	1.72%
$k-8$	77.82%	82.13%	4.31%
$k-9$	77.73%	79.62%	1.89%
$k-10$	77.73%	80.70%	2.97%



Gambar 8. Grafik Hasil Akurasi

Dari hasil akurasi pada Tabel 3. bisa disimpulkan bahwa nilai k dengan akurasi terbaik adalah pada $k-7$ (k -NN) dan $k-2$ (k -NNSB). Tetapi untuk perbandingan atau selisih hasil akurasi antara k -NN dengan k -NNSB adalah pada $k-1$, dimana $k-1$ pada k -NNSB mendapatkan hasil akurasi yang lebih baik daripada k -NN. Pada Gambar 8. menunjukkan bahwa setiap hasil akurasi yang dihasilkan oleh k -NNSB selalu lebih baik daripada k -NN pada setiap nilai $k-1$ sampai dengan $k-10$.

Pada Tabel 4. dan Tabel 5. melihat hasil terbaik pada k -NN yaitu pada $k-7$ dan k -NNSB pada $k-2$. Dapat dilihat bahwa di hampir semua

aspek kalau k -NNSB mendapatkan hasil yang lebih baik daripada k -NN.

Tabel 4. Tingkat Akurasi k -NN pada Nilai $k-7$

	true Registrasi	true Tidak Registrasi	class precision
pred. Registrasi	1654	413	80.02%
pred. Tidak Registrasi	109	213	66.15%
class recall	93.82%	34.03%	

Tabel 5. Tingkat Akurasi k -NNSB pada Nilai $k-2$

	true Registrasi	true Tidak Registrasi	class precision
pred. Registrasi	1670	147	91.91%
pred. Tidak Registrasi	122	450	78.67%
class recall	93.19%	75.38%	

3.1.4. Hasil Waktu Komputasi

Selain akurasi, penelitian yang dilakukan adalah menguji hasil dari waktu komputasinya. Pada Tabel 5. bisa dilihat hasil yang didapatkan k -NNSB selalu mendapatkan hasil yang lebih baik pada setiap nilai k yang sudah diujikan.

Tabel 6. Hasil Waktu Komputasi

Parameter k	Waktu Komputasi		Selisih
	k -NN	k -NN Sample Bootstraping	
$k-1$	3	3	0
$k-2$	3	3	0
$k-3$	3	3	0
$k-4$	3	3	0
$k-5$	3	3	0
$k-6$	3	3	0
$k-7$	3	3	0
$k-8$	3	3	0

Parameter k	Waktu Komputasi		Selisih
	k - N N	k -NN Sample Bootstrapping	
k -9	3	3	0
k -10	3	3	0

Dari hasil yang didapatkan seperti pada Tabel 6. maka bisa disimpulkan bahwa hasil waktu komputasi antara k -NN dengan k -NNSB tidak terdapat selisih atau perbedaan.

3.2. Pembahasan

Penelitian yang sudah dilakukan pada k -NN dan k -NNSB menunjukkan bahwa pada setiap nilai variabel k yang sudah diujikan, k -NNSB selalu mendapatkan hasil yang lebih baik dalam akurasi, sedangkan untuk waktu komputasi yang dihasilkan antara k -NN dengan k -NNSB membutuhkan waktu yang sama disetiap uji nilai k -1 sampai k -10 yaitu selama 3 detik.

Maka dengan ini k -NNSB menjadi lebih baik daripada k -NN untuk menangani masalah klasifikasi yang mana pada kasus ini menggunakan dataset PMB 2011-2013 STMIK Widya Pratama Pekalongan.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Dari penelitian yang dilakukan terhadap k -NN dan k -NNSB menggunakan aplikasi rapidminer 5.3.015. yang mana ini difungsikan untuk mengklasifikasikan data herregistrasi mahasiswa baru, menghasilkan akurasi yang terbaik adalah pada k -NNSB dengan nilai variabel k -2 yaitu sebesar 88.74%. Sedangkan untuk selisih terbesar yang dilakukan antara k -NN dan k -NNSB adalah pada nilai variabel k -1 yaitu sebesar 16.53% dimana k -NNSB lebih unggul dibanding k -NN.

Dapat disimpulkan bahwa untuk mengklasifikasikan data herregistrasi mahasiswa baru, k -NNSB bisa lebih unggul dibandingkan dengan menggunakan k -NN. Dimana kasus di penelitian ini menggunakan dataset PMB STMIK 2011-2013.

4.2. Saran

Untuk kedepannya didalam melakukan penelitian, khususnya dalam kasus didalam k -NN, ada beberapa hal yang mungkin bisa dilakukan yaitu:

1. Mencoba menggunakan dataset yang lebih bermacam-macam lagi, agar bisa lebih membuktikan hasil yang didapatkan.
2. Dilakukan resampling atau preprosesing data yang akan digunakan pada k -NN.
3. Dilakukannya pembobotan dengan menggunakan metode-metode pembobotan yang ada seperti Weighted PCA dan lain sebagainya.

5. Referensi

- Ann, Morrison, Oulasvirta Antti, and Peltonen Peter. "Like Bees Around the Hive: A Comparative Study of a Mobile Augmented Reality Map." *ACM (ACM)*, April 2009.
- Axel , Kupper. *Location-based Services : Fundamentals and Operation*. John Wiley & Sons Ltd, 2005.
- B. Efron, D. Rogosa, R. Tibshirani. "Resampling Methods of Estimation." *International Encyclopedia of the Social & Behavioral Sciences*, 2011: 13216–13220.
- Choonsung, Shin, Lee Wonwoo, and Suh Youngjung. "CAMAR 2.0: Future Direction of Context-Aware Mobile Augmented Reality." *IEEE (IEEE)*, 2009.
- Daniel, Wagner, and Schmalstieg Dieter. "History and Future of Tracking for Mobile Phone Augmented Reality." *IEEE (IEEE)*, 2009.

- . "History and Future of Tracking for Mobile Phone Augmented Reality." Gwangju, 2009. 7 - 10.
- Dawei, Liu, and Lee Moon-Chuen. "Mobile Localization in Outdoor Environments." *IEEE (IEEE)*, 2008.
- Georg, Klein, and Murray David. "Parallel Tracking and Mapping on a Camera Phone." *IEEE (IEEE)*, 2009.
- Gerhard, Reitmayr, and Drummond Tom W. "Initialisation for Visual Tracking in Urban Environments." *IEEE (IEEE)*, 2007.
- Gerhard, Reitmayr, Langlotz Tobias, and Wagner Daniel. "Simultaneous Localization and Mapping for Augmented Reality." *IEEE (IEEE)*, 2010.
- Gerhard, Schall, Wagner Daniel, and Reitmayr Gerhard. "Global Pose Estimation using Multi-Sensor Fusion for Outdoor Augmented Reality." *IEEE (IEEE)*, 2009.
- . "Global Pose Estimation using Multi-Sensor Fusion for Outdoor Augmented Reality." Washington, 2009. 153-162.
- Michael, Rohs, and Oulasvirta Antti. "Target Acquisition with Camera Phones when used as Magic Lenses." *ACM (ACM)*, 2008.
- Oliver, Bimber, and Raskar Ramesh. *Spatial Augmented Reality*. Mitsubishi Electric Research Laboratory, 2005.
- W.T., Fong, Ong S.K., and Nee A.Y.C. "A Differential GPS Carrier Phase Technique for Precision Outdoor AR Tracking." *IEEE (IEEE)*, 2008.
- . "A Differential GPS Carrier Phase Technique for Precision Outdoor AR Tracking." Cambridge, 2008. 25-28.
- Wikipedia. *Wikipedia*. Agustus 16, 2016. <https://id.wikipedia.org> (accessed Agustus 16, 2016).