

IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK PENENTUAN KELULUSAN MAHASISWA TEPAT WAKTU

¹Hamdani, ²Torkis Nasution

^{1,2}Teknik Informatika, STMIK Amik Riau,

Jl. Purwodadi Indah, Km. 10 Panam, Pekanbaru 28294

Email: hamdani@sar.ac.id, torkisnasution@sar.ac.id

ABSTRAK

Pelaksanaan seleksi merupakan upaya Perguruan Tinggi untuk mendapatkan calon mahasiswa yang berkualitas. Data hasil uji calon mahasiswa baru mampu mendeskripsikan kualitas akademik dan menghubungkan dengan lulus tepat waktu. Mengenali kualitas akademik mahasiswa diperlukan dalam pelaksanaan proses perkuliahan untuk memperoleh hasil yang optimal. Kondisi nyata saat ini, kelulusan tepat waktu belum tercapai secara optimal, perlu peningkatan untuk mencapai batas kewajaran. Data yang telah ada perlu dilakukan klasifikasi berdasarkan kualitas akademik, sehingga diperoleh prediksi kelulusan tepat waktu. Oleh karena itu, diajukan sebuah upaya untuk mengatasi masalah dengan menerapkan Algoritma K-Nearest Neighbor untuk meng-clustering data hasil ujian calon mahasiswa baru. Prosedur yang dilakukan adalah menentukan Jumlah cluster data, menentukan titik pusat cluster, menghitung jarak objek dengan centroid, mengelompokkan objek. Jika kelompok data hasil perhitungan baru sama dengan hasil perhitungan kelompok data baru maka selesailah perhitungannya. Data yang akan digunakan dalam clustering adalah hasil ujian masuk calon mahasiswa baru 3 tahun, dan telah dinyatakan diterima masuk di STMIK Amik Riau. Penelitian ini bertujuan untuk memprediksi kelulusan mahasiswa tepat waktu atau tidak. Hasil penelitian pada pengujian nilai k, akurasi maksimum didapatkan ketika k=5 yakni mencapai 99.25%. Akurasi akan menurun jika nilai k semakin besar maka hasil semakin tidak akurat.

Keywords: klasifikasi, lulus tepat waktu, kualitas akademik

1 PENDAHULUAN

Seleksi masuk ke STMIK Amik Riau dilakukan dengan memberikan soal-soal tes yang harus mereka kerjakan. Tes tersebut bertujuan untuk mengetahui kemampuan dan pengetahuan calon mahasiswa. Tes yang diberikan adalah tes potensial akademik (TPA), nilai tes Bahasa Inggris, dan Wawancara. Pelaksanaan ujian telah dilakukan dalam 10 tahun terakhir. Sampai saat ini penggunaan data tersebut masih terbatas pada seleksi masuk dan perangkingan mahasiswa yang dinyatakan lulus. Kelulusan merupakan satu diantara item penilaian dalam proses akreditasi institusi suatu perguruan tinggi. Elemen penilaian pada lulusan perguruan tinggi salah satunya adalah perguruan tinggi memiliki angka efisien edukasi yang ideal (BANPT, 2007). Kelulusan tepat waktu mahasiswa ditentukan oleh berbagai faktor, antara lain:

- a. Motivasi mahasiswa dalam menempuh pendidikan, walaupun dapat dipastikan bahwa setiap mahasiswa yang terdaftar menginginkan lulus namun tidak semua didorong oleh motif yang sama.
- b. Pelaksanaan proses pembelajaran,
- c. Status mahasiswa, umumnya mahasiswa yang mengambil kelas reguler pagi relatif akan lebih cepat selesai dibanding dengan kelas reguler malam.
- d. Sumber biaya, turut mempengaruhi terlebih lagi mahasiswa tersebut membiayai dirinya sendiri. Mahasiswa yang bekerja untuk membiayai kuliah relatif lebih lambat kelulusan, hal ini dipengaruhi oleh ritme kerja mahasiswa yang bersangkutan.

- e. Faktor kemampuan akademik, kondisi menjadi penentu karena kemampuan mahasiswa dalam menyelesaikan tugas mata kuliah didukung oleh kemampuan diri sendiri.

Dalam penelitian akan menggunakan data hasil ujian calon mahasiswa masuk ke STMIK Amik Riau, permasalahan adalah:

- a. Data hasil uji calon mahasiswa baru mendeskripsikan kualitas namun sampai saat ini belum pernah di ekspos ataupun digali;
- b. Data awal potensi belajar mahasiswa baru belum ada, padahal data tersebut dapat digali dari data hasil uji seleksi mahasiswa baru;
- c. Dalam upaya menggali potensi mahasiswa di dalam proses perkuliahan data akademik dan potensi diri diperlukan;
- d. Kelulusan tepat waktu merupakan indikator kemampuan perguruan tinggi dalam mengelola pelaksanaan akademik.

2 TINJAUAN PUSTAKA

Untuk menentukan posisi penelitian ini dengan penelitian yang terdahulu maka dibuatlah State Of The Art. Pada penelitian ini membandingkan dengan tiga jurnal yang berkaitan dengan Implementasi Algoritma K-Nearest Neighbor untuk Penentuan Kelulusan Mahasiswa Tepat Waktu. Pada tahun 2019, Sekar Rizkya Rani dan kawan-kawan dalam penelitiannya yang berjudul “Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Kelulusan Siswa pada SMK Anak Bangsa” yang diterbitkan pada Prosiding Seminar Nasional Riset Information Science (SENARIS). Melakukan analisis terhadap aspek sikap dan kehadiran siswa untuk memprediksi kelulusan tepat waktu. Penelitian ini bertujuan untuk menghasilkan kelulusan tepat waktu yang didasarkan pada penilaian sikap dan penilaian kehadiran disekolah. Hasil penelitian yang diperoleh adalah nilai $K=5$ dengan tingkat akurasi sebesar 93,55 % yang ditetapkan sebagai K-Optimal. Nilai $K=5$ diterapkan pada algoritma K-Nearest Neighbor untuk prediksi kelulusan siswa berdasarkan kehadiran, sikap dan nilai pengetahuan. Pada tahun 2018, eneliti Eri Sasmita Susanto dan kawan-kawan dengan judul “Prediksi Kelulusan Mahasiswa Magister Teknik Informatika Universitas Amikom Yogyakarta Menggunakan Metode K-Nearest Neighbor” yang diterbitkan pada Jurnal Teknologi Informasi. Penelitian fokus untuk mengetahui uji kelayakan prediksi kelulusan mahasiswa Universitas AMIKOM Yogyakarta. Data yang diolah bersifat numerik dan tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar. Input dari sistem ini adalah Data sampel berupa data mahasiswa tahun 2014-2015. pengujian pada penelitian ini menggunakan dua pengujian yaitu data testing dan data training. Kriteria yang digunakan dalam penelitian ini adalah, IP Semester 1-4, capaian SKS, Status Kelulusan. Output dari sistem ini berupa hasil prediksi kelulusan mahasiswa yang terbagi menjadi

dua yaitu tepat waktu dan kelulusan tidak tepat waktu. Hasil pengujian menunjukkan bahwa Berdasarkan penerapan $k=14$ dan $k\text{-fold}=5$ menghasilkan performa yang terbaik dalam memprediksi kelulusan mahasiswa menggunakan indeks prestasi 4 semester dengan nilai akurasi= 98,46%, precision= 99.53% dan recall =97.64%.

Pada tahun 2019, Lalu Abd Rahman Hakim dan kawan-kawan melakukan penelitian dengan judul Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN). Dasar penelitian ini adalah presentasi naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi kampus. Berdasarkan data dari Bagian Prodi di 3 tahun terakhir presentasi kelulusan mahasiswa hanya 25% mahasiswa dari total yang keseluruhan mahasiswa yang dapat menyelesaikan studinya dengan tepat waktu. Pada penelitian ini menggunakan algoritma K-Nearest Neighbor yang bertujuan untuk dapat mengidentifikasi kelulusan mahasiswa pada kasus baru dengan cara mengadaptasi solusi dari kasus sebelumnya yang memiliki kedekatan dengan kasus baru. Algoritma ini berperan untuk mendapatkan nilai kedekatan kasus baru terhadap kasus lama, yang selanjutnya populasi terbanyak pada area K dengan nilai terdekat didapatkan mahasiswa tersebut di prediksi apakah lulus tepat waktu atau tidak tepat waktu. Penelitian ini menggunakan metode waterfall Roger S. Pressman yaitu

Communication, Planning, Modelling, dan Construction. Berdasarkan pengujian yang dilakukan menggunakan K-Fold Cross Validation didapatkan akurasi tertinggi pada model yang ketiga sebesar 80% ketika k-fold ke-4 dan 61% ketika nilai K=1. Sedangkan pengujian menggunakan Confusion Matrix didapatkan akurasi tertinggi sebesar 98% pada K=1 untuk klasifikasi “Tepat Waktu”, dan 98% pada K=2 untuk klasifikasi “Tidak Tepat Waktu”.

Pada tahun 2018, Arie Yandi Saputra dan Yogi Primadasa melakukan penelitian dengan judul “Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour” yang diterbitkan pada jurnal Techno.COM. dasar penelitian ini adalah Bertambahnya mahasiswa tiap tahunnya menyebabkan penumpukan data mahasiswa. Prediksi ketepatan kelulusan mahasiswa dirancang untuk mendukung prodi untuk membimbing mahasiswa agar kelulusan tepat waktu. Menurut peneliti, dengan mengetahui prediksi status kelulusan mahasiswa berjalannya perkuliahan, maka prodi di bantu pembimbing akademik dapat memberi perhatian khusus terhadap mahasiswa yang di prediksi tidak lulus tepat waktu (terlambat) sehingga mahasiswa tersebut dapat memperbaiki indeks prestasinya tiap semester agar dapat lulus tepat waktu. Penelitian ini menggunakan 9 data training dan 1 data testing. Dengan menggunakan k=5 yang diterapkan menggunakan metode K-Nearest Neighbor untuk prediksi kelulusan mahasiswa.

3 METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini konsentrasi pada pendeteksi penyimpangan maupun pelanggaran disiplin oleh mahasiswa dengan berlandaskan pada peraturan yang sudah ada pada STMIK Amik Riau maupun Kemenristek-Dikti. Kerangka kerja dibangun sedemikian rupa sehingga dapat digunakan sebagai acuan untuk mencapai lulus tepat waktu. Berikut ini prosedur dalam rancangan penelitian :

1. Data dikumpulkan, data hasil ujian calon mahasiswa baru di kumpulkan yang meliputi ujian Tes Potensi Akademik, Tes Bahasa Inggris, dan Tes Wawancara.
2. Semua data yang dikumpulkan disimpan dalam database yang dapat di load oleh sistem maupun pimpinan.
3. Data diambil dari database kemudian dilakukan pengolahan data pre processing dan disimpan kembali.
4. Data yang dibaca akan diukur validitas dan reliabilitas berdasarkan aturan akademik, disiplin kemahasiswaan, non akademik, kepatutan dalam organisasi, faktor internal mahasiswa yang bersangkutan, dan kompilasi data.
5. Data direkam kembali ke database sebagai data hasil kompilasi.
6. Data yang ada di olah menggunakan metode K-NN.
7. Melakukan pengujian atas hasil metoda K-NN, dengan menggunakan data sample.
8. Memutuskan hasil capaian mahasiswa lulus tepat waktu.
9. Membuat kesimpulan dan implikasi.

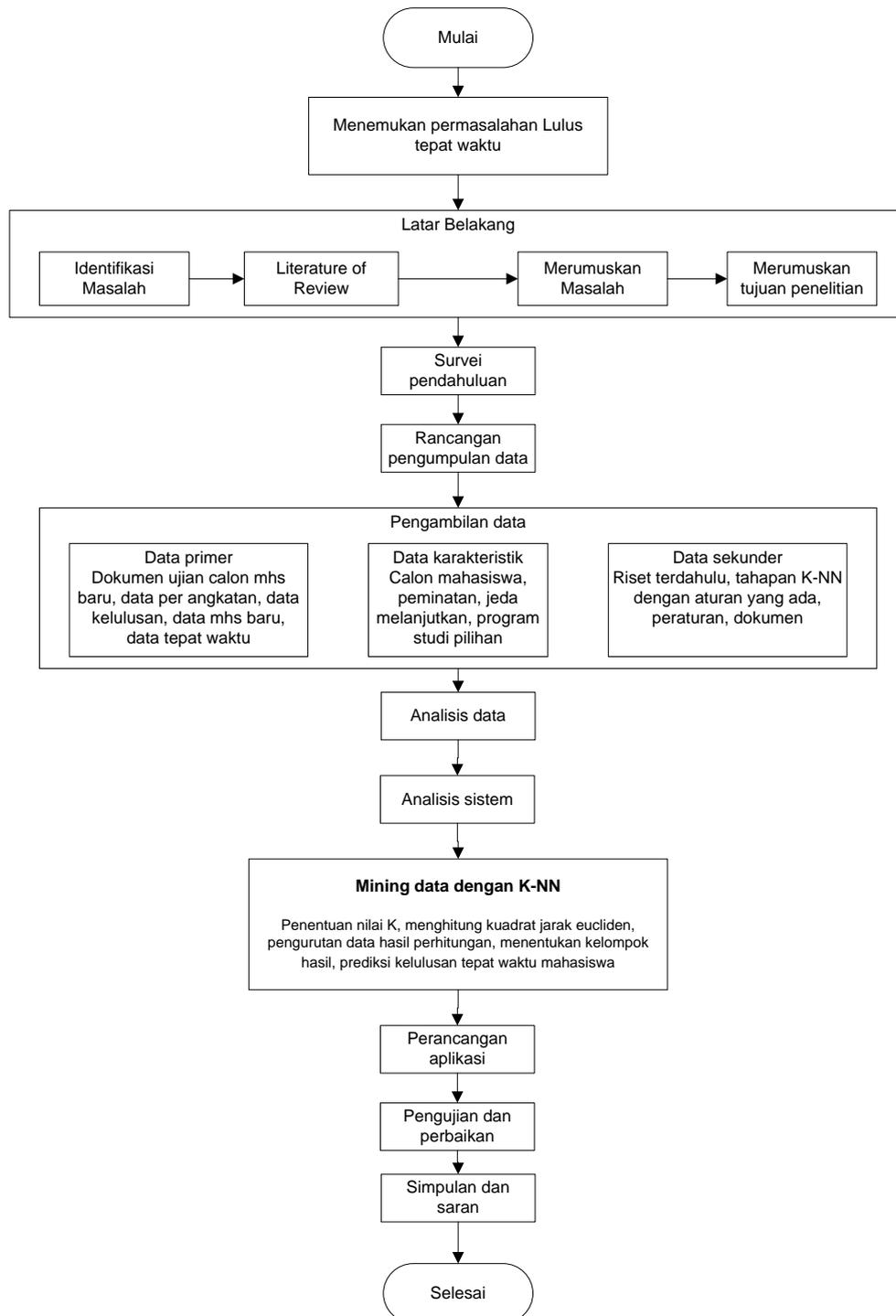
Semua hasil pada setiap siklus proses direkam pada database system, sehingga dapat disajikan informasi yang lengkap untuk setiap siklus. Adapun teknik yang digunakan dalam penelitian ini adalah:

1. Analisa Prosedur, melakukan analisa terhadap prosedur yang telah ditetapkan STMIK Amik Riau terhadap standar pelaksanaan ujian dengan demikian data yang diperoleh sah.
2. Pengamatan Dokumen, melakukan amatan terhadap dokumen yang sudah tersedia berhubungan dengan data hasil ujian, daftar peserta, pelaksanaan ujian, dan sifat ujian.
3. Sampling, melakukan pengambilan *sample* data terhadap populasi, sehingga dapat ditarik kesimpulan terhadap hasil prediksi sistem berbanding dengan data sample yang ada.
4. Studi pustaka, menemukan artikel ilmiah yang relevan untuk mendapat riset terdahulu tentang prediksi lulus tepat waktu, keterhubungan ujian calon mahasiswa baru dengan lulus tepat waktu;

Data yang diperoleh dijadikan untuk melakukan pengujian dan pembuktian penerapan dengan metode K-NN.

3.3. Bagan Alir Penelitian

Bagan alir penelitian yang diuraikan ke dalam masing-masing tahapan, dapat dilihat pada gambar dibawah ini.



Gambar 1. Bagan Alir Penelitian

Tahapan perancangan penelitian diatas, dilakukan secara berurutan dan berkesinambungan antara satu tahap dengan tahap lainnya. Selama proses penelitian ini berlangsung peneliti melakukan pencatatan kegiatan pada *logbook* yang menggambarkan kemajuan penelitian.

3.4 Analisa Data

Dalam data kualitatif data berbentuk hasil analisis menggunakan kata-kata bukan berupa angka tetapi hasil diperoleh dari pengamatan dilapangan. Analisis kualitatif digunakan pada data yang diperoleh dari hasil observasi tentang kebiasaan karyawan dalam mengambil foto serta mendokumentasikan. Dalam pengumpulan data dari dua sudut yaitu dari tenaga pendidik dan tenaga kependidikan. Data tersebut diolah dan dianalisis untuk perencanaan pembelajaran berikutnya. Dalam melakukan analisis data harus berdasarkan prosedur dan langkah-langkah tertentu. Berikut ini adalah beberapa langkah-langkah dalam analisis data:

1. Pengumpulan Data, tahap awal dari aktivitas analisis data adalah pengumpulan data yang akan dianalisis.
2. Tahap Perbaikan, yaitu proses pemeriksaan kejelasan dan kelengkapan terkait pengisian instrumen pengumpulan data.
3. Tahap Interpretasi, yaitu proses identifikasi dan klasifikasi terhadap semua pernyataan yang ada pada instrumen pengumpulan data berdasarkan variable yang sedang diteliti.
4. Tahap Pengujian, yaitu proses pengujian kualitas data, baik dari sisi validitas dan realibilitas instrumen dari pengumpulan data.
5. Tahap Mendeskripsikan Data, yaitu proses membuat deskripsi data dengan menyajikannya dalam bentuk tabel frekuensi atau diagram dengan beragam ukuran tendensi sentral maupun ukuran dispersi. Tujuannya adalah agar memahami karakteristik data sampel dari suatu penelitian.
6. Tahap Pengujian Hipotesis, yaitu proses pengujian terhadap proposisi apakah bisa diterima atau ditolak, apakah memiliki makna atau tidak. Berdasarkan tahap inilah nantinya akan dibuat kesimpulan atau keputusan.

4 HASIL DAN PEMBAHASAN

4.1. Hasil

Setelah melakukan analisa terhadap perancangan dengan tahapan data mining untuk menghasilkan prediksi tepat waktu dengan menggunakan algoritma K-Nearest Neighbor. Analisa ini berakhir dengan melakukan proses data mining yang sesungguhnya, maka hasil yang dicapai oleh peneliti adalah untuk mengetahui prediksi lulus tepat waktu berdasarkan hasil ujian seleksi masuk perguruan tinggi. Peneliti menggunakan aplikasi RapidMiner untuk melakukan proses data mining, yang menghasilkan informasi prediksi lulus tepat waktu.

4.1.2 Analisis Klasifikasi K-NN

Penganalisis untuk klasifikasi data pada penelitian ini menggunakan algoritma klasifikasi K-Nearest Neighbor. Algoritma KNN bekerja berdasarkan jarak terpendek dari *training* dan *test* dalam penentuannya, kemudian mengambil mayoritas hasil ketentuan yang telah didapatkan untuk dijadikan prediksi dari *test*. Sebagai pengimplementasian algoritma dalam aplikasi, maka diambil dari 3 data *training* yang dijadikan sebagai nilai *k*. Pada data *training* tersebut telah memiliki klasifikasi masing-masing, yaitu data hasil ujian *training* 2 dan 4 mempunyai klasifikasi pengulangan, sedangkan pada data hasil ujian *training* ke 3 mempunyai klasifikasi pemilihan.

4.4. Data Selection

Data yang digunakan dalam penelitian ini, yaitu data penerimaan calon mahasiswa berdasarkan hasil ujian 3 tahun terakhir. Data berasal dari tahun akademik 2016/2017, 2017/2018, 2018/2019. Kemudian data tersebut diseleksi dan digunakan untuk diolah dalam memprediksi mahasiswa lulus tepat waktu. Adapun atribut yang digunakan dalam penentuan prediksi mahasiswa lulus tepat waktu adalah nama calon mahasiswa, hasil ujian tertulis, hasil ujian wawancara, predikat. Data training dapat dilihat pada tabel 1.

Tabel 1 Tabel Data Selection

No	IdUjian	CalonMahasiswa	Kelas	UjiBhs Inggris	Uji Kecak Logika	Nilai Huruf	Hasil Akhir
1	802-0-43762	Azza Nurleli Fuadiah	Siang	42.00	68.00	A	Lulus
2	802-0-32530	Winda Indi Astuti	Siang	48.00	58.00	B	Lulus
3	802-0-59763	Hazraf Padmadani	Siang	36.00	58.00	B	Lulus
4	802-0-29216	Raja Kevin Alhafid	Siang	44.00	58.00	B	Lulus
5	802-0-89302	Imam Anugrah Adha	Siang	38.00	48.00	B	Lulus
6	802-0-57021	Wilda Adriani Frastika	Siang	30.00	46.00	B	Lulus
7	802-0-30149	Arif Budi Santoso	Siang	50.00	54.00	B	Lulus
8	802-0-53932	Wahyu Pradana	Siang	36.00	54.00	A	Lulus
9	802-0-83596	Rafly Muhammad Fitra	Siang	38.00	56.00	B	Lulus
10	802-0-98367	Prima Sanggul	Siang	38.00	56.00	B	Lulus
11	802-0-98368	Ridho Wandira	Siang	30.00	36.00	D	Lulus
12	802-0-98369	Rafita	Siang	45.00	56.00	D	Lulus
...
444	802-804-80370	Roy Daniel Nainggolan	Siang	0	0.00	E	Tidak Lulus

4.5. Preprocessing

Tahap preprocessing dilakukan untuk pengelompokan jenis ujian berdasarkan jumlah penerimaan mahasiswa baru untuk mempermudah dalam proses perhitungan prediksi. Setelah data terkelompokkan lalu semua dijumlahkan sehingga menjadi data penerimaan mahasiswa baru. Dapat dilihat pada tabel 2.

Tabel 2 Data Preprocessing

Calon Mahasiswa	UjiBhs Inggris	Uji Kecak Logika	Hasil Akhir
Azza Nurleli Fuadiah	42.00	68.00	Lulus
Winda Indi Astuti	48.00	58.00	Lulus
Hazraf Padmadani	36.00	58.00	Lulus
Raja Kevin Alhafid	44.00	58.00	Lulus
Imam Anugrah Adha	38.00	48.00	Lulus
Wilda Adriani Frastika	30.00	46.00	Lulus
Arif Budi Santoso	50.00	54.00	Lulus
Wahyu Pradana	36.00	54.00	Lulus
Rafly Muhammad Fitra	38.00	56.00	Lulus
Prima Sanggul	38.00	56.00	Lulus
Ridho Wandira	30.00	36.00	Lulus
Rafita	45.00	56.00	Lulus
..
Roy Daniel Nainggolan	0	0.00	Tidak Lulus

Berdasarkan tabel 3 data training yang digunakan pada penelitian ini hanya diambil dari 2 tahun sebelumnya yaitu penerimaan calon mahasiswa tahun akademik 2017/2018 dan 2018/2019 yang menjadi data training.

Tabel 3. Data Uji Lulus Tepat Waktu

Calon Mahasiswa	UjiBhs Inggris	Uji Kecak Logika	Hasil Akhir
Azza Nurleli Fuadiah	42.00	68.00	Lulus
Winda Indi Astuti	48.00	58.00	Lulus
Hazraf Padmadani	36.00	58.00	Lulus
Raja Kevin Alhafid	44.00	58.00	Lulus
Imam Anugrah Adha	38.00	48.00	Lulus
Wilda Adriani Frastika	30.00	46.00	Lulus
Arif Budi Santoso	50.00	54.00	Lulus
Wahyu Pradana	36.00	54.00	Lulus
Rafly Muhammad Fitra	38.00	56.00	Lulus
Prima Sanggul	38.00	56.00	Lulus
Ridho Wandira	30.00	36.00	Lulus
Rafita	45.00	56.00	Lulus
..
Roy Daniel Nainggolan	0	0.00	Tidak Lulus

4.6. Transformation

Pada tahap transformation ini hasil dari pengelompokan data preprocessing kemudian digunakan untuk data training. Proses pembentukan data training berdasarkan data yang ada, data harus di seleksi terlebih dahulu untuk menentukan atribut mana yang dapat mempengaruhi kelulusan tepat waktu yang disebut data target, dimana data target merupakan data yang berisikan atribut yang akan menjadi atribut yang relevan dan mendukung dalam proses data mining. Adapun data training yang digunakan adalah seperti yang ditunjukkan pada tabel 4.

Tabel 4 Data Training

Calon Mahasiswa	UjiBhs Inggris	Uji Kecak Logika	Hasil Akhir
Azza Nurleli Fuadiah	42.00	68.00	Lulus
Winda Indi Astuti	48.00	58.00	Lulus
Hazraf Padmadani	36.00	58.00	Lulus
Raja Kevin Alhafid	44.00	58.00	Lulus
Imam Anugrah Adha	38.00	48.00	Lulus
Wilda Adriani Frastika	30.00	46.00	Lulus
Arif Budi Santoso	50.00	54.00	Lulus
Wahyu Pradana	36.00	54.00	Lulus
Rafly Muhammad Fitra	38.00	56.00	Lulus
Prima Sanggul	38.00	56.00	Lulus
Ridho Wandira	30.00	36.00	Lulus
Rafita	45.00	56.00	Lulus
..
Roy Daniel Nainggolan	0	0.00	Tidak Lulus

Berdasarkan tabel 5 data training yang digunakan pada penelitian ini hanya diambil dari 2 tahun sebelumnya yaitu penerimaan calon mahasiswa tahun akademik 2017/2018 dan 2018/2019 yang menjadi data training. Data training dikelompokkan menjadi 2, yaitu data input dan target. Dimana data input merupakan data penerimaan mahasiswa pada gelombang 1, sedangkan data target menggunakan data gelombang 2. Selanjutnya untuk gelombang 3 digunakan sbagai data input dengan target keluaran pada gelombang 1 tahun akademik 2018/2019, dan seterusnya. Sedangkan data testing yang digunakan untuk prediksi disajikan pada table abc

Tabel 5 Data Testing

CalonMahasiswa	UjiBhsInggris	UjiKecakapanLogika	Hasil Akhir
Nurmahdiyah	85.00	78.00	Lulus
Mega Susanti	75.00	70.00	Lulus
Muhammad Ibnu Rushandy	86.00	80.00	Lulus
Aida Nisa	74.00	82.00	Lulus
Krida Pratama	88.00	90.00	Lulus
Winda Krisnayanti	0.00	0.00	Lulus
Rizki Yeputra	80.00	86.00	Lulus
Sayyid Alpharizki	75.00	90.00	Lulus
Muhammad Qorib Alqowiy	82.00	86.00	Lulus

4.7. K-Nearest Neighbor

Berdasarkan dari tahapan data mining untuk algoritma K-Nearest Neighbor, adapun langkah-langkah adalah:

1. Penentuan nilai k, penentuan nilai k yang digunakan tidak memiliki aturan yang baku, namun pada penelitian ini nilai k yang diguanakn adalah 5.
2. Menghitung kuadrat jarak eucliden (query-instance) masing-masing objek terhadap sampel data yang diberikan. Koordinat query-instance adalah (50,40) dimana nilai tersebut berasal dari nilai attribut yang akan diproduksi.
3. Pengurutan data hasil perhitungan, jarak yang telah didapatkan kemudian diurutkan dari yang paling dekat jaraknya sampai yang apling jauh (asending).
4. Menentukan kelompok data hasil uji berdasarkan label mayoritas dari k tetangga terdekat. Karena nilai k=5 maka diambil maka diambil 5 jarak terkecil
5. Dengan menggunakan K-NN yang paling mayoritas maka dapat diprediksikan mahasiswa lulus tepat waktu.

4.7.1. Penentuan Nilai K

Jumlah data tetangga terdekat ditentukan dan dinyatakan dengan k. Penentuan nilai k terbaik dapat ditentukan dengan optimasi parameter, misalnya dengan menggunakan

4.7.2. Menghitung Euclidean distance

Menggunakan persamaan

$$D(x, y) = \sqrt{\sum_{k=1}^n (X_{training} - Y_{testing})^2}$$

Berikut hasil dari euclidean distance masing-masing objek terhadap data testing yang diberikan

Tabel 6 Euclidean distance data testing

Calon Mahasiswa	UjiBhsInggris	UjiKecakapanLogika	Eucliden
Azza Nurleli Fuadiah	42	68	29.12
Winda Indi Astuti	48	58	18.11
Hazraf Padmadani	36	58	22.80
Raja Kevin Alhafid	44	58	18.97
Imam Anugrah Adha	38	48	14.42
Wilda Adriani Frastika	30	46	20.88
Arif Budi Santoso	50	54	14.00
Wahyu Pradana	36	54	19.80
Rafly Muhammad Fitra	38	56	20.00
Prima Sanggul	38	56	20.00
Ridho Wandira	30	36	20.40
Rafita	45	56	16.76

Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclidian terkecil. Setelah objek-objek tersebut di urutkan maka menentukan peringkat dari data yang mempunyai jarak terkecil ke terbesar, hasil perngurutan peringkat jarak dapat di lihat pada tabel 7 di bawah ini :

Tabel 7 Pengurutan Peringkat Jarak

Calon Mahasiswa	UjiBhsInggris	UjiKecakapanLogika	Eucliden	Urutan
Azza Nurleli Fuadiah	42	68	29.12	12
Winda Indi Astuti	48	58	18.11	9
Hazraf Padmadani	36	58	22.80	10
Raja Kevin Alhafid	44	58	18.97	11
Imam Anugrah Adha	38	48	14.42	1
Wilda Adriani Frastika	30	46	20.88	8
Arif Budi Santoso	50	54	14.00	3
Wahyu Pradana	36	54	19.80	6
Rafly Muhammad Fitra	38	56	20.00	2
Prima Sanggul	38	56	20.00	7
Ridho Wandira	30	36	20.40	5
Rafita	45	56	16.76	4

Dengan menggunakan kategori mayoritas,maka dapat hasil klasifikasi Setelah mengumpulkan kategori Y dengan K =5 maka langkah selanjutnya adalah mengumpulkan kategori mayoritas. Hasil dari kategori mayoritas dapat di lihat pada tabel 8 di bawah ini:

Tabel 8 Hasil Kategori Mayoritas

Calon Mahasiswa	Uji Bhs Inggris	Uji Kecakapan Logika	Eucliden	Urutan	Apakah termasuk nearest neighbor (K)
Azza Nurleli Fuadiah	42	68	29.12	12	Tidak
Winda Indi Astuti	48	58	18.11	9	Tidak
Hazraf Padmadani	36	58	22.80	10	Tidak
Raja Kevin Alhafid	44	58	18.97	11	Tidak
Imam Anugrah Adha	38	48	14.42	1	Ya
Wilda Adriani Frastika	30	46	20.88	8	Tidak
Arif Budi Santoso	50	54	14.00	3	Ya
Wahyu Pradana	36	54	19.80	6	Tidak
Rafly Muhammad Fitra	38	56	20.00	2	Ya
Prima Sanggul	38	56	20.00	7	Tidak
Ridho Wandira	30	36	20.40	5	Tidak
Rafita	45	56	16.76	4	Ya

1.7.3. Data Hasil Uji

Berdasarkan label mayoritas dari k tetangga terdekat. Karena nilai k=60 maka diambil maka diambil 3 jarak terkecil. Dari hasil klasifikasi diatas jadi, data testing termasuk kategori Tepat Waktu dikarenakan jumlah mayoritas Tepat Waktu berjumlah 4 dan Terlambat berjumlah 1. Hasil data testing dapat dilihat pada tabel 9

Tabel 9 Hasil Data Testing

Calon Mahasiswa	Uji Bhs Inggris	Uji Kecakapan Logika	Eucliden	Urutan	Apakah termasuk nearest neighbor (K)	Y = kategori nearest neighbor
Azza Nurleli Fuadiah	42	68	29.12	12	Tidak	-
Winda Indi Astuti	48	58	18.11	9	Tidak	-
Hazraf Padmadani	36	58	22.80	10	Tidak	-
Raja Kevin Alhafid	44	58	18.97	11	Tidak	-
Imam Anugrah Adha	38	48	14.42	1	Ya	Baik
Wilda Adriani Frastika	30	46	20.88	8	Tidak	-
Arif Budi Santoso	50	54	14.00	3	Ya	Kurang
Wahyu Pradana	36	54	19.80	6	Tidak	-
Rafly Muhammad Fitra	38	56	20.00	2	Ya	Cukup
Prima Sanggul	38	56	20.00	7	Tidak	-
Ridho Wandira	30	36	20.40	5	Tidak	-
Rafita	45	56	16.76	4	Ya	Jelek
Winda Krisnayanti	0	0	0	13	Tidak	-

Dengan menggunakan kategori nearest neighbor yang paling mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung. Mempunyai kualitas 2 Jelek dan 2 baik, karena nilai Y sama maka yang dipilih adalah jelek. Maka kita simpulkan bahwa prediksi kelulusan tepat waktu dengan menetapkan hasil nilai attribute $X_1=50$, $X_2=3$ termasuk ketegori jelek. Dari penjelasan diatas algoritma K-Nearest ini memiliki kelebihan yang terdiri dari :

1. Dapat menghasilkan data yang kuat atau jelas (khususnya jika menggunakan turunan perkalian kuadrat pada besaran jarak)

2. Efektif jika digunakan untuk data yang besar

Dari beberapa kelebihannya maka K-Nearest juga memiliki kekurangan yaitu :

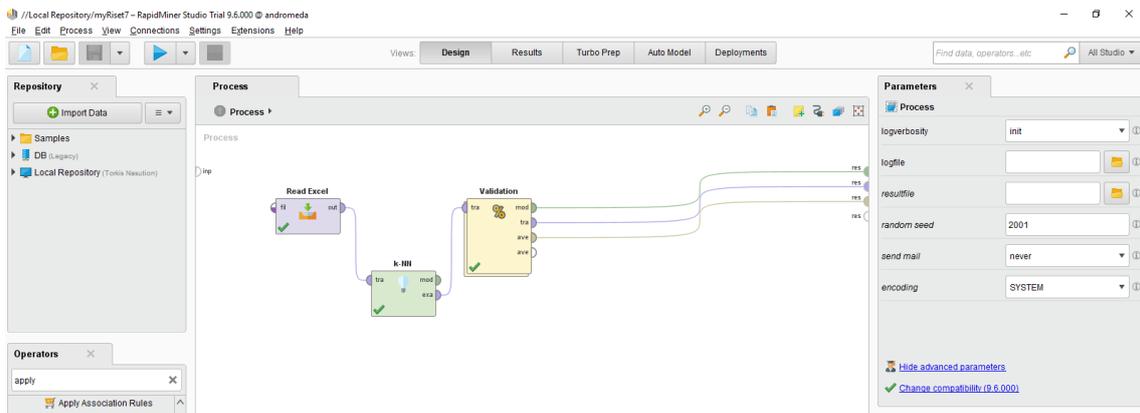
1. Membutuhkan nilai K sebagai parameter.

2. Jarak dari data percobaan tidak dapat jelas dengan tipe jarak yang digunakan dan dengan atribut yang digunakan untuk menghasilkan hasil yang terbaik, maka harus menggunakan semua atribut atau hanya 1 aribut yang telah pasti.

3. Perhitungan harga sangat tinggi karena percobaan ini membutuhkan perhitungan jarak dari beberapa query untuk semua data percobaan.

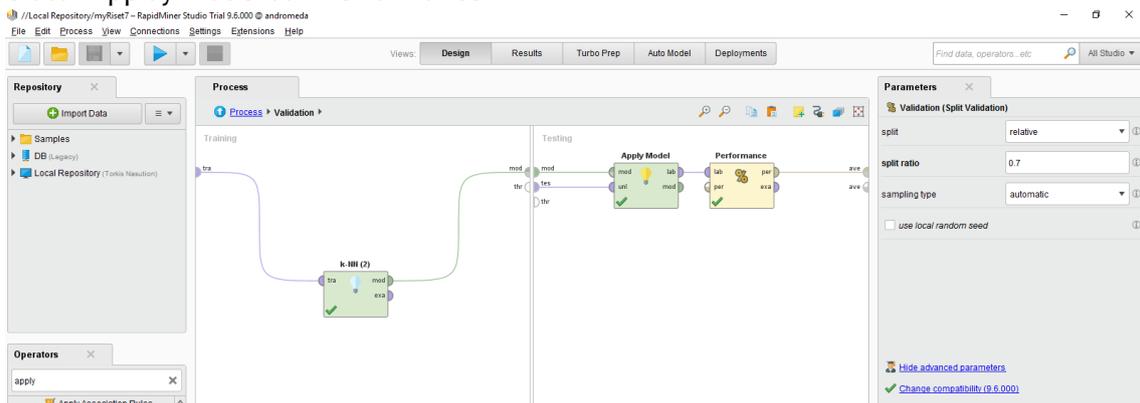
1.7.4. Implementasi Rapid Miner

RapidMiner merupakan salah satu software data mining pengolahan data set untuk mnecari pola data sesuai dengan tujuan dari pengolahan data tersebut. Tidak semua algoritma yang dapat sesuai atau dapat mengolah data set yang ada, harus dilakukan penyesuaian pola data dan sesuai dengan tujuan dari pengolahan data tersebut. Pada proses terdapat 3 operator masing-masing adalah: Pembacaan data eksternal dari Microsoft Excel Read Excel, K-NN, dan Validation, ditunjukkan pada gambar 4.1



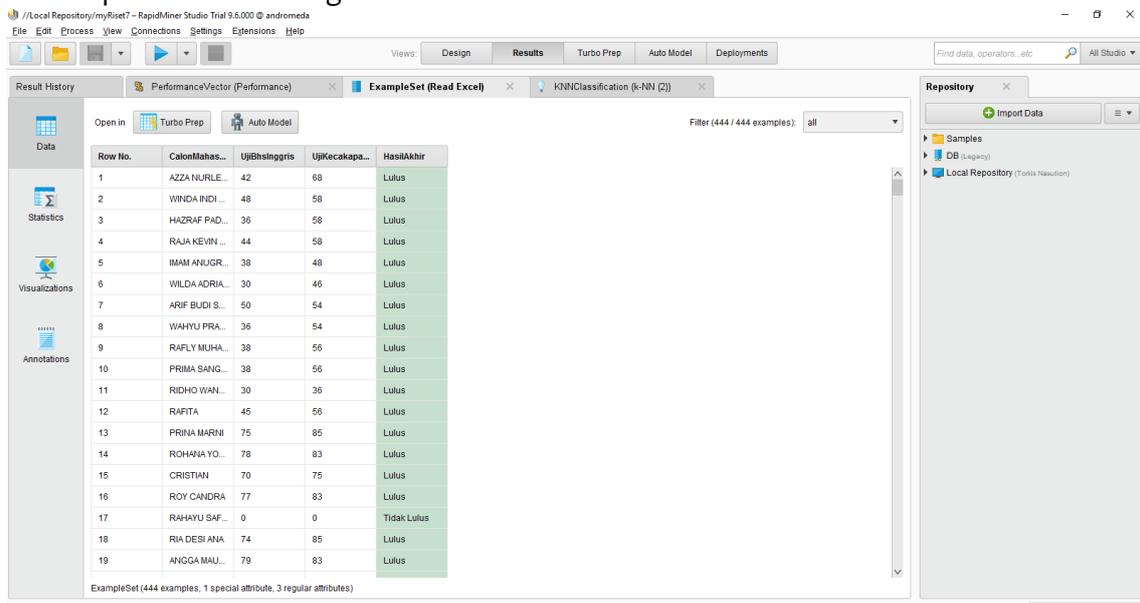
Gambar 2 Model Proses Classification K-Nearest Neighbor

Operator validation memiliki sub model yang di bagi menjadi 2, bagian data training dan bagian data testing. Pada data training memiliki operator K-NN, pada bagian data testing memiliki operator Apply Model dan Performance



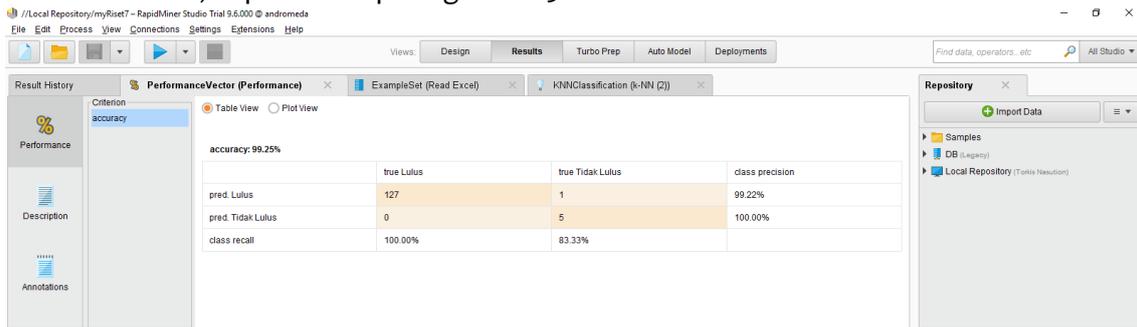
Gambar 3 Model Sub Validation

Gambar 4 membuktikan data telah dibaca dengan baik dan seluruhnya oleh RapidMiner. Data ini akan di proses untuk menghasilkan klasifikasi KNN



Gambar 4 Hasil Pembacaan Data oleh RapidMiner

Tingkat hasil akurasi dari performance vector sebesar 99.25% untuk proses yang dilaksanakan pada model diatas, dapat dilihat pada gambar 5.



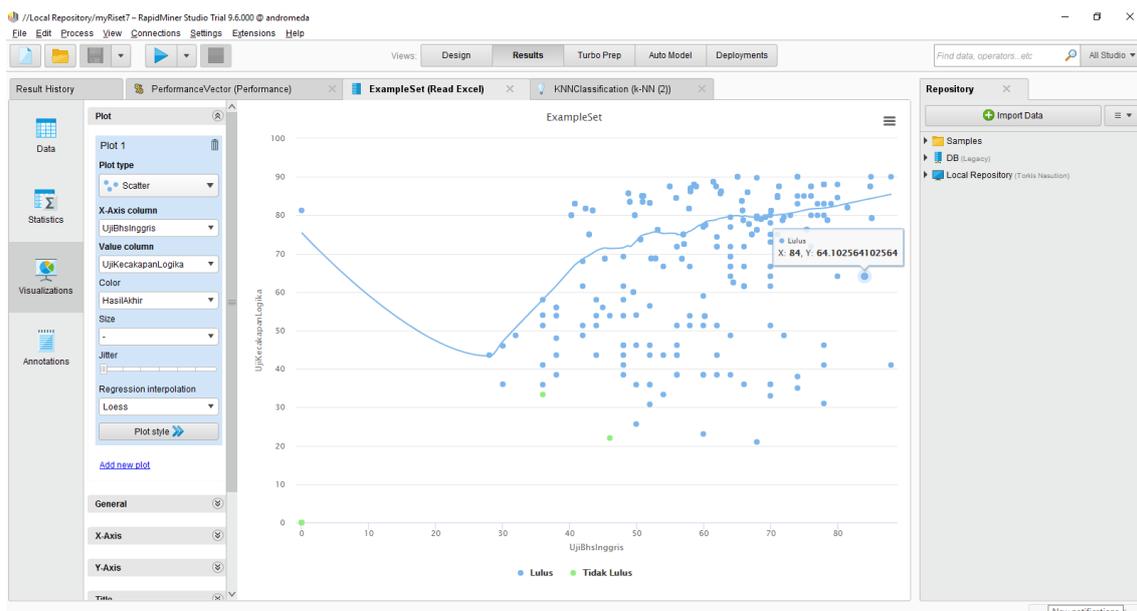
Gambar 5 Hasil perhitungan nilai akurasi

Gambar 6 Rapid miner melakukan klasifikasi terhadap data, dan menghasilkan KKN Clasification pada lulus dan tidak lulus tepat waktu



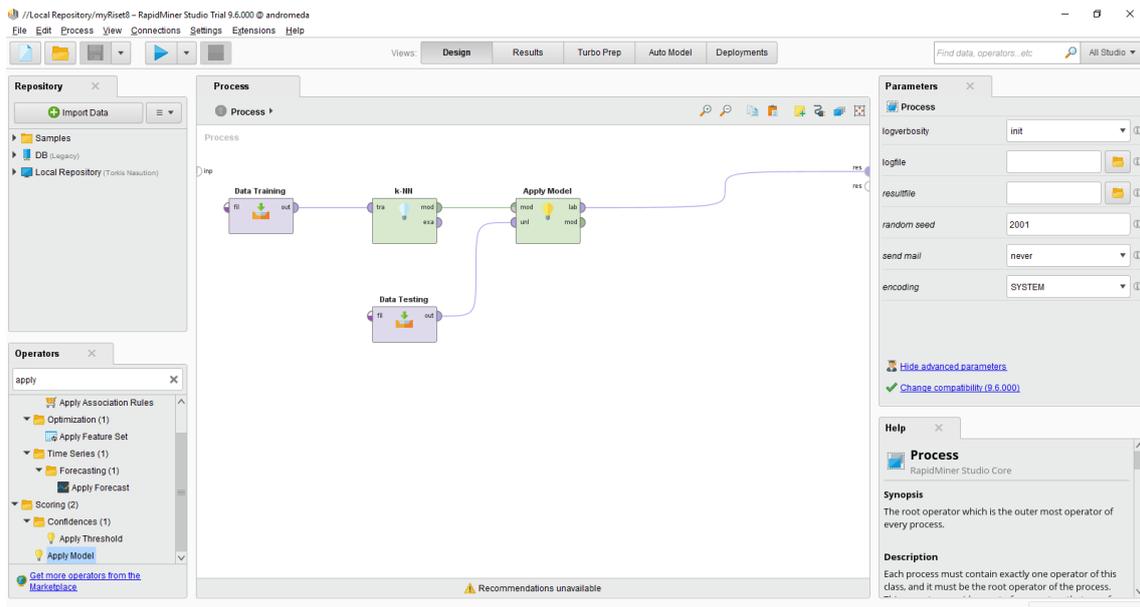
Gambar 6 Model Klasifikasi K-Nearest Neighbor

Gambar 7 menjelaskan secara grafik tampilan sebaran kelulusan tepat waktu dan tidak lulus tepat waktu.



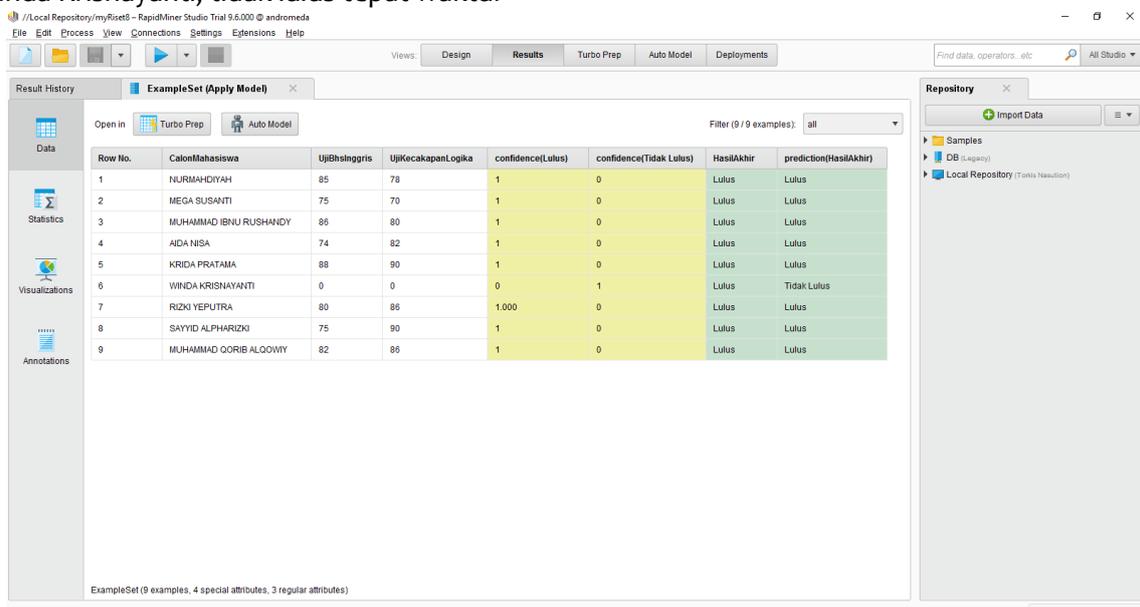
Gambar 7 Sebaran data kelulusan tepat waktu

Gambar 8 rancangan model untuk melakukan pengujian terhadap data testing berdasarkan data training yang telah di ujia terlebih dahulu tingkat akurasiya.



Gambar 8 Model Sub Data Testing berdasarkan Data Trining

Gambar 9 mendeskripsikan bahwa hasil pengujian kelulusan tepat waktu atas nama nama Winda Krisnayanti, tidak lulus tepat waktu.



Gambar 9 Hasil pengujian

5 KESIMPULAN

Setelah mendeskripsikan hasil dan pembahasahan, dapat ditarik kesimpulan berupa hasil penelitian adalah:

1. Pada pengujian nilai k, akurasi maksimum didapatkan ketika k=5 yakni mencapai 99.25%. Akurasi akan menurun jika nilai k semakin besar.
2. Akurasi akan meningkat jika jumlah data semakin banyak. Pada pengujian komposisi data latih, akurasi terbaik yang diperoleh mencapai 84%
3. Komposisi data latih berpengaruh terhadap akurasi karena saat komposisinya seimbang, maka data latih merata di setiap kelas data uji

Dalam hal ini penulis ingin memberikan beberapa saran yang mungkin berguna untuk pengembangan lebih lanjut penelitian ini antara lain :

Hamdani, Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Kelulusan Mahasiswa Tepat Waktu

1. Dapat dikembangkan untuk pembuatan aplikasi sehingga pelaksanaan klasifikasi data lebih fleksibel.
2. Menambah variabel data, untuk menambah akurasi prediksi.
3. Melaksanakan pengujian klasifikasi secara tuntas waktu, untuk untuk memperoleh data secara realtime.

REFERENSI

- [1] Muhamad Ichsan, Meyti Eka Apriani, Sandi Prasetyaningsih, 2016, ANALISIS DAN IMPLEMENTASI CLOUD RECOGNITION DAN DEVICE STORAGE PADA AUGMENTED REALITY JENIS BUAH, Jurnal Ilmiah Komputer dan Informatika (KOMPUTA), Vol. 5, No. 1, Maret 2016, ISSN : 2089-9033
- [2] Tomy Hidayat, Fitri Susanti, Yahdi Siradj, 2017, IMPLEMENTASI LAYANAN PRIVATE CLOUD STORAGE MENGGUNAKAN OWN CLOUD DAN MONITORING DENGAN ZENOSS, SSN : 2442-5826 e-Proceeding of Applied Science : Vol.3, No.1 April 2017
- [3] Muhamad Dany Kurniawan, Ibnu Irvan Hanafi, Thera Frista Dewi Karina Bulan, Rico Agung Firmansyah, 2016, DESIGN DAN IMPLEMENTASI CLOUD STORAGE BERBASIS WEB PADA RT/RW NET MAJU JAYA, Seminar Nasional Teknologi Informasi dan Multimedia 2016, STMIK AMIKOM Yogyakarta, 6-7 Februari 2016, ISSN : 2302-3805
- [4] Reggy Lintang Perdana, Heru Supriyono, IMPLEMENTASI CLOUD STORAGE DI KANTOR KECAMATAN NGEMPLAK BOYOLALI, J urnal Emitor Vol.17 No. 01 ISSN 1411-8890
- [5] Irfan Santiko, Rahman Rosidi, Seta Agung Wibawa, 2017, PEMANFAATAN PRIVATE CLOUD STORAGE SEBAGAI MEDIA PENYIMPANAN DATA E-LEARNING PADA LEMBAGA PENDIDIKAN JURNAL TEKNIK INFORMATIKA VOL.10 NO.2, 2017, p-ISSN 1979-9160, e-ISSN 2549-7901