

Komparasi Teknik *Decision Tree*, KNN dan *Naïve Bayes* dalam Analisis Klasifikasi Narapidana Kasus *Narkotika* pada LAPAS

Herdian Ananta Putra dan Sigit Widiyanto

Jurusan Manajemen Sistem Informasi, Program Pascasarjana, Universitas Gunadarma

Jl. Margonda Raya No.100 Pondok Cina Depok Jawa Barat

E-mail : herdianputra89@gmail.com , sigitwidiyanto@staff.gunadarma.ac.id

Abstrak

Narapidana yang ada di dalam lapas akan menempati sel atau kamar yang telah disediakan oleh pihak lapas. Namun saat ini pada lapas belum ada klasifikasi narapidana terutama kasus narkotika berdasarkan tingkat risiko untuk penempatan kamar. Hal ini akan berdampak terhadap adanya potensi gangguan keamanan dan ketertiban terutama peredaran dan pengendalian narkotika dari dalam lapas. Dampak lainnya yang mungkin terjadi adalah narapidana dengan kasus narkotika dapat mempengaruhi narapidana lain untuk melakukan kejahatan serupa dari dalam lapas. Fakta yang sering muncul bahwa pelaku kejahatan narkotika dapat digolongkan sebagai bandar, pengedar dan pengguna. Tujuan penelitian ini adalah untuk menentukan teknik klasifikasi yang paling tepat untuk digunakan dalam klasifikasi narapidana kasus narkotika dengan membandingkan 3 teknik klasifikasi yaitu *Decision tree*, KNN, dan *Naïve Bayes*. Hasil penelitian menemukan bahwa teknik *Decision tree* dengan nilai *accuracy* 98,86%, sedangkan untuk teknik kNN dengan nilai *accuracy* 92,57% dan teknik *Naïve Bayes* dengan nilai *accuracy* 68,43%. Dengan hasil tersebut maka teknik *Decision Tree* merupakan teknik yang paling tepat digunakan berdasarkan nilai *accuracy*.

Kata kunci : lapas, narapidana, narkotika, klasifikasi, *decision tree*.

Pendahuluan

Sebagai salah satu bagian dari sistem peradilan pidana yang ada Indonesia, Lembaga Pemasyarakatan (Lapas) memiliki peran dalam pelaksanaan pembinaan narapidana dan anak didik pemasyarakatan melalui sistem pemasyarakatan berbasis pendidikan, rehabilitasi dan reintegrasi. Lembaga Pemasyarakatan sebagai ujung tombak pelaksanaan asas pengayoman merupakan tempat untuk mencapai tujuan tersebut diatas melalui pendidikan, rehabilitasi dan reintegrasi (Setneg, 1995, p.29). Berdasarkan data pada akhir Desember 2019, jumlah Warga Binaan Pemasyarakatan yang ada di lapas/rutan wilayah DKI Jakarta berjumlah 18.045 orang dengan jumlah kapasitas daya tampung idealnya hanya 5.791 orang (Ditjen PAS, 2019). Artinya dengan kondisi ini ada kelebihan daya tampung sebanyak 12.254 atau mengalami kelebihan kapasitas 312%. Sedangkan pegawai/petugas pemasyarakatan di lapas/rutan DKI Jakarta berjumlah 1.779 orang dengan 1.161 orang diantaranya bertugas sebagai satuan pengamanan (Ditjen PAS, 2019). Artinya satu orang petugas

pengamanan mengawasi 16 WBP (1:16).

Kondisi demikian tidak ideal dimana yang dihadapi oleh petugas pemasyarakatan ini adalah orang-orang yang melakukan tindak pidana yang berpotensi mencelakai petugas. Salah satu tindak pidana yang banyak ada di dalam lapas/rutan adalah narkotika. Fakta pahit yang harus diterima adalah peredaran gelap narkotika dikendalikan dari dalam lapas/rutan[1]. Maka tidak salah jika kejahatan narkotika dikategorikan sebagai kejahatan luar biasa (*extra ordinary crime*) atau kejahatan lintas batas negara (*transnational crime*)[2]. Narapidana ini bahkan dapat mempengaruhi petugas lapas untuk menyediakan beberapa fasilitas yang dibutuhkan untuk pengendalian barang haram tersebut[3]. Narapidana tersebut dapat mengendalikan bisnis barang haramnya melalui telepon seluler maupun telepon satelit untuk komunikasi dengan bandar maupun pengedar yang ada diluar lapas.

Klasifikasi narapidana dengan kasus narkotika yang tidak tepat dapat memunculkan kegiatan pengendalian bisnis narkotika dari dalam lapas. Jika bandar dan pengedar dijadikan dalam satu ka-

mar maka berpotensi dapat melakukan pengendalian barang haram tersebut. Pada lapas penempatan kamar narapidana masih belum berdasarkan klasifikasi narapidana kasus narkoba berdasarkan tingkat risiko. Sehingga dikhawatirkan akan menimbulkan potensi peredaran gelap narkoba di dalam lapas. Tentunya dengan kondisi seperti itu program pembinaan narapidana tidak dapat berjalan maksimal.

Berdasarkan referensi dokumen yang penulis telaah, masing-masing status narapidana kasus narkoba tersebut mempunyai risiko yang berbeda-beda[4]. Masalah yang dapat diidentifikasi dari kondisi tersebut adalah belum adanya klasifikasi berdasarkan tingkat risiko. Pada penulisan ini, penulis ingin menemukan teknik klasifikasi narapidana kasus narkoba yang tepat pada lapas menggunakan proses data mining dengan beberapa teknik klasifikasi.

Metode Penelitian

Penelitian ini adalah merupakan pemodelan untuk klasifikasi narapidana kasus narkoba yang tepat pada lapas. Penelitian ini menggunakan beberapa variabel yang terkait dengan hasil penelitian yang telah dilakukan pada penjara atau lapas di Indonesia dan luar negeri. Hasil dari klasifikasi narapidana kasus narkoba ini adalah dapat memprediksi apakah narapidana tersebut memiliki risiko tinggi atau risiko rendah. Dengan klasifikasi narapidana yang tepat diharapkan dapat menekan potensi peredaran gelap narkoba dari dalam lapas. Pada penelitian ini menggunakan proses *data mining* dengan metode klasifikasi. Penulis menggunakan 3 teknik klasifikasi yaitu *Decision tree*, KNN dan *Naïve Bayes*. Tujuan penelitian ini membandingkan 3 teknik klasifikasi tersebut untuk menemukan teknik yang paling tepat digunakan pada proses mining data narapidana kasus narkoba berdasarkan nilai *accuracy* yang dihasilkan pada proses mining.

Data mining

Data mining merupakan sebuah proses penggalian data dan informasi dan belum diketahui sebelumnya namun data dan informasi tersebut dapat dipahami dan berguna. Data tersebut berasal dari suatu database yang besar dan dapat digunakan untuk pengambilan keputusan. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan. Data mining memungkinkan pengguna menemukan sebuah pengetahuan dalam data suatu *database*, lihat Gambar 1.

Pengertian *Data mining* menurut beberapa ahli sebagai berikut:

1. Menurut Davies (2004) data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi [5];
2. Menurut Han (2006) Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu – ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti neural network, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* [6];
3. Menurut Retno Tri Vlandari (2017) dalam bukunya *Data mining Teori dan Aplikasi Rapidminer*, *data mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstrasi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. *Data mining* digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases (KDD)* [7].

Proses dari data mining itu mempunyai beberapa tujuan sebagai berikut :

1. *Explanatory*, sebagai sarana untuk menjelaskan suatu kondisi atau keadaan dari hasil penelitian;
2. *Confirmatory*, sebagai sarana untuk mengklarifikasi sebuah pernyataan;
3. *Exploratory*, sebagai sarana untuk mencari pola baru yang belum ditemukan.

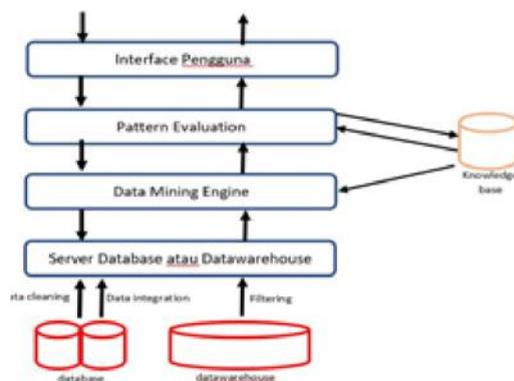
Tahapan melakukan data mining adalah sebagai berikut :

1. *Data cleaning* merupakan tahap menghilangkan data yang tidak konsisten dan *noise*);
2. *Data integration* (penggabungan data dari sumber yang berbeda);
3. *Data selection* merupakan tahap pemilihan data yang relevan dengan analisis akan dikembalikan ke dalam *database*;
4. *Data transformation* (data diubah menjadi bentuk yang sesuai);
5. *Data mining* merupakan tahap proses untuk mengekstrak sebuah pola data);

6. *Pattern evolution* merupakan tahap untuk mengidentifikasi pola yang bernilai dan menarik);
7. *Knowledge presentation* (di mana gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan kepada user).

Arsitektur utama dari data mining, terdiri dari beberapa elemen sebagai berikut:

1. *Database, data warehouse*, media penyimpanan terdiri dari satu atau beberapa *database*. Perlu dilakukan pembersihan dan integrasi dari data tersebut. *Database, data warehouse* melakukan pencarian data yang sesuai dengan kebutuhan pengguna;
2. *Knowledge base*, pengetahuan yang digunakan untuk panduan dalam pencarian pola data;
3. *Data mining engine*, merupakan elemen penting dan idealnya terdiri dari kumpulan modul dan fungsi yang digunakan dalam proses penentuan karakteristik, klasifikasi dan kluster yang kemudian dijalankan berdasarkan algoritma yang ada;
4. *Pattern evaluation*, berinteraksi dengan modul-modul yang ada pada data mining dan berfungsi untuk menemukan pola yang ada dalam suatu database kemudian diolah sehingga mendapatkan pengetahuan yang sesuai;
5. *Interface pengguna*, merupakan modul komunikasi dan interaksi antara pengguna dengan sistem untuk menentukan proses dari data mining itu sendiri.



Gambar 1: Arsitektur *Data mining*

Klasifikasi

Klasifikasi adalah salah satu metode paling umum digunakan dalam data mining dan sudah banyak

diimplementasikan dalam kehidupan sehari-hari. Klasifikasi adalah proses untuk menemukan model yang menjelaskan atau membedakan kelas data dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri berupa aturan “*if-then*” berupa *Decision tree*, *neural network* atau formula matematis. Klasifikasi juga merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui. Klasifikasi dapat digambarkan sebagai berikut. *Data input*, disebut juga *training set* terdiri atas banyak *record* yang masing-masing memiliki beberapa atribut. Kemudian, tiap *record* diberi sebuah label *class* khusus. Tujuannya untuk menganalisa data input dan mengembangkan deskripsi atau model akurat untuk tiap *class* menggunakan fitur-fitur pada data. Deskripsi *class* ini digunakan untuk mengklasifikasikan data pengujian lainnya dengan label *class* tidak diketahui. Deskripsi tersebut juga dapat digunakan untuk memahami tiap *class* dalam data.

Decision tree

Decision tree adalah salah satu metode klasifikasi yang paling banyak digunakan, karena mudah untuk diimplementasikan oleh manusia. *Decision tree* adalah model prediksi yang menggunakan *hierarchy structure* atau *tree structure*. Konsep dari pohon keputusan adalah mengubah data menjadi *Decision tree* dengan beberapa aturan-aturan keputusan yang dibuat. Manfaat dari penggunaan *Decision tree* adalah dapat membreak *down* proses pengambilan keputusan yang kompleks menjadi lebih *simple*, sehingga pengambil keputusan nantinya akan lebih menginterpretasikan solusi dari permasalahan yang ada. *Decision tree* juga dapat bermanfaat untuk eksplorasi data, menemukan hubungan yang tersembunyi antar variabel input dengan variabel target. *Decision tree* menggabungkan antara eksplorasi data dengan pemodelan, sehingga cocok untuk proses awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain. *Decision tree* juga mampu menghilangkan perhitungan atau data-data yang tidak diperlukan. Sebab, sampel yang ada biasanya hanya diuji berdasarkan kriteria atau kelas tertentu saja.

Naïve Bayes

Algoritma *Naïve bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naïve bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema*

Bayes. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. *Naïve Bayes* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas. Menurut Olson Dellen (2008) menjelaskan *Naïve Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek[8]. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dr " *master* " tabel keputusan.

K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek, berdasarkan k buah data latih yang jaraknya paling dekat dengan objek tersebut. Syarat nilai k adalah tidak boleh lebih besar dari jumlah data latih, dan nilai k harus ganjil dan lebih dari satu. Dekat atau jauhnya jarak data latih yang paling dekat dengan objek yang akan diklasifikasi dapat dihitung dengan menggunakan metode *cosine similarity*. *Cosine similarity* merupakan salah satu cara atau metode yang dapat digunakan untuk melihat sejauh mana kemiripan isi antar dokumen. Dalam hal ini *cosine similarity* berfungsi untuk menguji ukuran yang dapat digunakan sebagai interpretasi kedekatan jarak berdasarkan kemiripan dokumen.

K-Nearest Neighbor (KNN) termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok k objek dalam *data training* yang paling dekat (mirip) dengan objek pada data baru atau *data testing* [9,10]. Contoh kasus, misal diinginkan untuk mencari solusi terhadap masalah seorang pasien baru dengan menggunakan solusi dari pasien lama. Untuk mencari solusi dari pasien baru tersebut digunakan kedekatan dengan kasus pasien lama, solusi dari kasus lama yang memiliki kedekatan dengan kasus baru digunakan sebagai solusinya.

Sumber data

Sumber data yang diperoleh peneliti adalah data primer dan data sekunder. Data primer diambil dari wawancara dengan salah satu pejabat di Direktorat Jenderal Pemasyarakatan. Data sekunder penulis peroleh dari Aplikasi Sistem *Database* Pemasyarakatan milik Direktorat Jenderal Pemasyarakatan. Sistem *Database* Pemasyarakatan

(SDP) adalah sistem mekanisme pelaporan dan konsolidasi pengelolaan data Warga Binaan Pemasyarakatan (WBP), yang berfungsi sebagai alat bantu kerja sesuai kebutuhan Unit Pelaksana Teknis (UPT) Pemasyarakatan, Kantor Wilayah dan Ditjenpas. Pada SDP ini berisi data warga binaan pemasyarakatan, kapasitas dan isi penghuni pada UPT lapas/rutan dan lain-lain. Data tersebut merupakan data narapidana kasus narkoba dari Lapas. Penulis tidak menyebutkan secara spesifik nama Lapas untuk melindungi kerahasiaan institusi lapas. Penulis mengambil tempat penelitian pada Lapas karena di Lapas tersebut tingkat overkapasitasnya paling tinggi di kota tersebut. Selain itu, pada lapas narapidana dengan kasus narkoba paling banyak mendominasi dibanding narapidana kasus lainnya

Pada Lapas terdapat jumlah warga binaan pemasyarakatan sebanyak 3412 orang (data per 30 Juli 2020) dengan berbagai macam kasus kejahatan. Diantara 3412 narapidana pada lapas, 1800 narapidana diantaranya adalah kasus narkoba. Maka hanya 1800 narapidana ini yang akan dimasukkan dalam penelitian ini. Data warga binaan pemasyarakatan ini berisi atribut nama (disamarkan), no registrasi, alamat, tempat tanggal lahir, lama pidana, usia, jenis kelamin, pekerjaan, lokasi blok/kamar, jenis suku, pendidikan dan pekerjaan. Pada Tabel 1 dan 2 dilampirkan contoh data warga binaan pemasyarakatan.

Tabel 1: Contoh Data Narapidana (1)

No Reg Instansi	Nama	Propinsi	Kota	Tempat Lahir	Tanggal Lahir
B.1.3.0043/D.MDN-19	NAPI 1	Aceh	Kab. Aceh Tamiang	Kab. Aceh Tamiang	1980-12-30
MT.0001/D-03	NAPI 2	Riau	Indragiri Hilir	Bone	1962-01-01
B.1.2.0349/D.PDG-28	NAPI 3	Sumatera Barat	Kepulauan Mentawai	Kepulauan Mentawai	1970-11-11
B.1.2.0351/DMRS-18	NAPI 4	Sumatera Barat	Dharmasraya	Nias Selatan	1976-03-27
B.1.3.0104/DTPG-20	NAPI 5	Kepulauan Riau	Lingga	Lain-lain	1981-07-15

Tabel 2: Contoh Data Narapidana (2)

Usia	Jenis Kelamin	Tingkat Pendidikan	Lama Pidana (tahun)	Lama Pidana (bulan)	Jenis Kejahatan	Pekerjaan
39	Laki-laki	SD	16	0	Narkoba	Buruh Nelayan/Perikanan
58	Laki-laki	Tidak Sekolah	0	0	Pembunuhan	Wiraswasta
49	Laki-laki	Tidak lulus SD	7	6	Narkoba	Buruh Tani/Perkebunan
44	Laki-laki	Tidak lulus SD	13	0	Perlindungan Anak	Karyawan Swasta
39	Laki-laki	SMA	12	6	Narkoba	Wiraswasta

Kebutuhan Perangkat

Pada implementasi ini penulis menggunakan perangkat keras (hardware) dan perangkat lunak (software) yang mendukung dalam penelitian ini. Perangkat keras terdiri dari sebuah notebook/laptop dengan spesifikasi :

1. *Processor* : Intel Core i7-8850 CPU @ 2,60GHz (12CPUs);
2. *Memory* : 16 GB;
3. *Hardisk* : SSD 256 GB;

4. *Operating System* : Windows 10 Pro 64 Bit.
5. Perangkat lunak yang digunakan adalah aplikasi *data mining* yaitu *RapidMiner*.

Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data dari Sistem Database Pemasarakatan (SDP) yang bersifat *private* dan *web based*. Penulis

diberikan ijin oleh pengampu data dalam hal ini Direktur Teknologi Informasi dan Kerja Sama Direktorat Jenderal Pemasarakatan untuk akses data dimaksud dengan menyamakan *record* yang bersifat pribadi seperti nama narapidana termasuk nama instansi yaitu lapas itu sendiri. *Software* yang digunakan untuk mengakses SDP adalah browser *Mozilla Firefox ver 89.02* dengan jaringan *private* (VPN) dari Direktorat Jenderal Pemasarakatan, lihat Gambar 2.



Gambar 2: Tampilan Aplikasi Sistem *Database* Pemasarakatan

Seleksi Data

Seleksi data adalah proses menganalisis data-data yang relevan dari database karena sering penulis temukan bahwa tidak semua data dibutuhkan dalam proses *data mining*. Data tersebut dipilih dan diseleksi dari database untuk di analisis. Sumber data yang digunakan dalam penelitian ini berasal dari data narapidana yang masih aktif dengan kasus narkotika pada lapas.

Dari semua data yang digunakan hanya nama narapidana (disamarkan), usia, kategori usia, lama pidana(dalam tahun), jenis kejahatan, jenis kejahatan narkotika, residivis dan kategori narapidana. Karena informasi yang ada didalamnya sudah mewakili informasi yang dibutuhkan untuk penelitian.

Pembersihan Data

Pada tahap ini dilakukan pembersihan data. Pembersihan data dilakukan agar tidak ada yang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan data seperti kesalahan cetak, sehingga data tersebut dapat dilakukan

proses *data mining*. Pembersihan data dilakukan setelah data tersebut di *export* ke dalam *file excel*. Tabel 3,4 adalah data sebelum dibersihkan dan setelah dilakukan pembersihan data.

Tabel 3: Tabel Data Sebelum Dilakukan Pembersihan Data

Nama	Usia	Kategori Usia	Lama Pidana (tahun)	Jenis Kejahatan	Residivis	Jenis Kejahatan Narkotika	Kategori Narapidana
NAPI 1	26	Dewasa	20	Narkotika	Tidak	Bandar, Pengedar	Resiko Tinggi
NAPI 2	25	Dewasa	20	Narkotika	Tidak	Bandar, Pengedar	Resiko Tinggi
NAPI 3	26	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 4	46	Dewasa	22	Narkotika	Ya	Bandar, Pengedar	Resiko Tinggi
NAPI 5	34	Dewasa	15	Narkotika	Tidak	Bandar, Pengguna	Resiko Tinggi

Tabel 4: Tabel Data Setelah Dilakukan Pembersihan Data

Nama	Usia	Kategori Usia	Lama Pidana (tahun)	Jenis Kejahatan	Residivis	Jenis Kejahatan Narkotika	Kategori Narapidana
NAPI 1	26	Dewasa	20	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 2	25	Dewasa	20	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 3	26	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 4	46	Dewasa	22	Narkotika	Ya	Bandar	Resiko Tinggi
NAPI 5	34	Dewasa	15	Narkotika	Tidak	Bandar	Resiko Tinggi

Transfomasi Data

Pada tahap ini dilakukan transformasi data dimana pada tahap ini mengubah data menjadi bentuk yang sesuai untuk diproses dalam data mining. Data narapidana ini ketika di export dari SDP sudah dalam bentuk *file excel (.xls)* dan aplikasi *RapidMiner* sudah mendukung data dengan format *excel (.xls maupun .xlsx)*. Dan perubahan atribut hanya pada nama atribut Kategori menjadi Kategori Narapidana.

Tabel 5: . Tabel Data training

Nama	Usia	Kategori Usia	Lama Pidana (tahun)	Jenis Kejahatan	Residivis	Jenis Kejahatan Narkotika	Kategori Narapidana
NAPI 1	26	Dewasa	20	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 2	25	Dewasa	20	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 3	26	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 4	46	Dewasa	22	Narkotika	Ya	Bandar	Resiko Tinggi
NAPI 5	34	Dewasa	15	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 6	38	Dewasa	12	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 7	49	Dewasa	19	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 8	35	Dewasa	11	Narkotika	Tidak	Bandar	Resiko Tinggi
NAPI 9	27	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 10	34	Dewasa	20	Narkotika	Tidak	Bandar	Resiko Tinggi

Pengujian Dengan Aplikasi Rapid-Miner

Pada tahap ini penulis melakukan pengujian *data training* dan *data testing*. Atribut data Kategori Usia dan Kategori Narapidana dianalisis menggunakan aplikasi *RapidMiner*. Data narapidana tersebut memiliki 1817 *record data*. Data yang akan diujikan dibagi menjadi dua bagian yaitu *training* dan

testing. Data training sejumlah 1090 data dan data testing 727 data. Tabel 5 dan 6 adalah data narapidana dengan format excel (.xls) yang akan diolah menggunakan aplikasi *RapidMiner*.

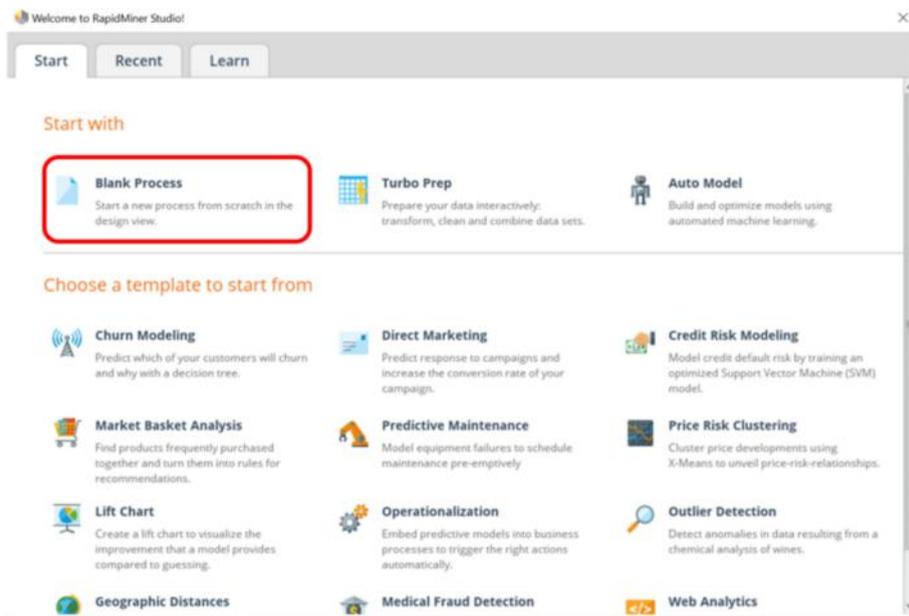
Tabel 6: Tabel Data testing

Nama	Usia	Kategori Usia	Lama Pidana (tahun)	Jenis Kejahatan	Residivis	Jenis Kejahatan Narkotika	Kategori Narapidana
NAPI 1	45	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 2	57	Dewasa	7	Narkotika	Ya	Pengedar	Resiko Tinggi
NAPI 3	32	Dewasa	5	Narkotika	Tidak	Pengedar	Resiko Rendah
NAPI 4	38	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 5	37	Dewasa	5	Narkotika	Ya	Pengguna	Resiko Rendah
NAPI 6	22	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 7	43	Dewasa	13	Narkotika	Ya	Bandar	Resiko Tinggi
NAPI 8	21	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 9	23	Dewasa	5	Narkotika	Tidak	Pengguna	Resiko Rendah
NAPI 10	50	Dewasa	6	Narkotika	Ya	Pengedar	Resiko Tinggi

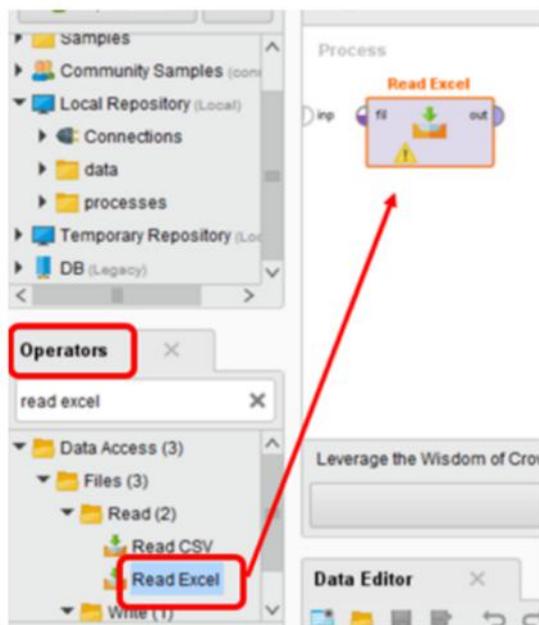
Yang pertama dilakukan adalah membuka aplikasi *RapidMiner* dan membuat *Blank Process* seperti Gambar 3.

Kemudian pada *menu Operator*, cari operator Read Excel kemudian drag dan drop ke *Windows Process*. Setelah itu lakukan lakukan *Import Configuration Data* seperti Gambar 4.

Setelah muncul form Data Import Wizard 1, lakukan import data excel yang akan dijadikan sebagai data training. Kemudian pada form Data Import Wizard 2 pilih semua data yang ada dalam file excel tersebut. Kemudian pada form Data Import Wizard 3 lalu pilih atribut target atau label yang digunakan yaitu atribut Kategori Narapidana, lihat Gambar 5. Setelah ditentukan label yang dipilih maka klik Finish.



Gambar 3: . Tampilan Blank Process Aplikasi Rapid Miner



Gambar 4: Drag dan Drop operator Read Excel

Format your columns.

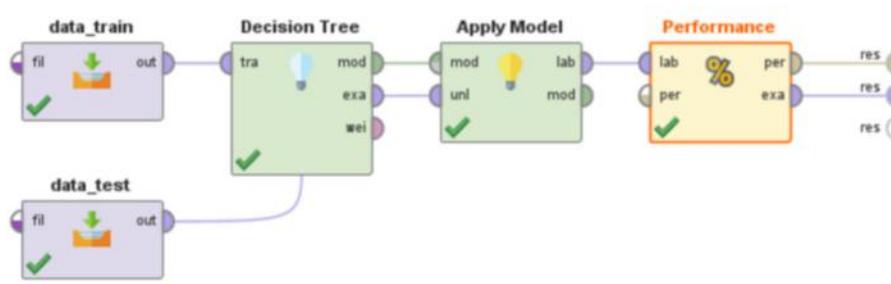
Replace errors with missing values

Kategori_U...	Lama_Pida...	Jenis_Keja...	Residivis	Jenis_Keja...	Kategori N...
polynomial	integer	polynomial	polynomial	polynomial	polynomial label
Dewasa	20	Narkoba	Tidak	Bandar	Risiko Tinggi
Dewasa	20	Narkoba	Tidak	Bandar	Risiko Tinggi

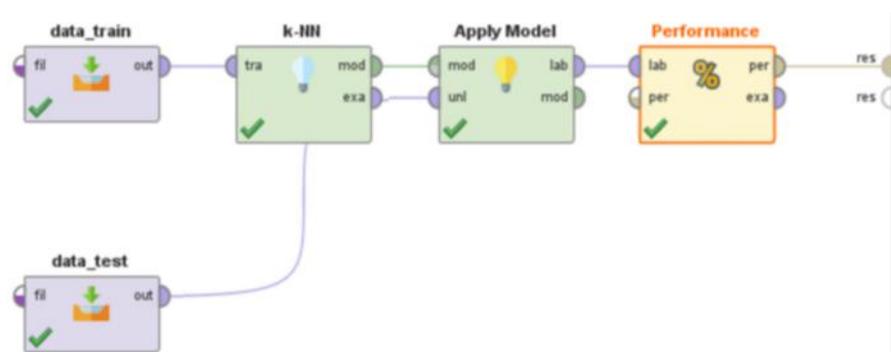
Gambar 5: Ubah Atribut Kategori Narapidana menjadi Label

Setelah dilakukan import data training dan data testing kemudian tambahkan operator Decision tree, Apply Model, Performance seperti Gambar 6. Setelah semuanya terhubung, klik ikon Run pada toolbar, untuk menampilkan hasil. Tunggu beberapa saat, komputer memerlukan waktu untuk menyelesaikan perhitungan. Waktu yang dibutuhkan komputer untuk menyelesaikan penghitungan tergantung dari spesifikasi komputer.

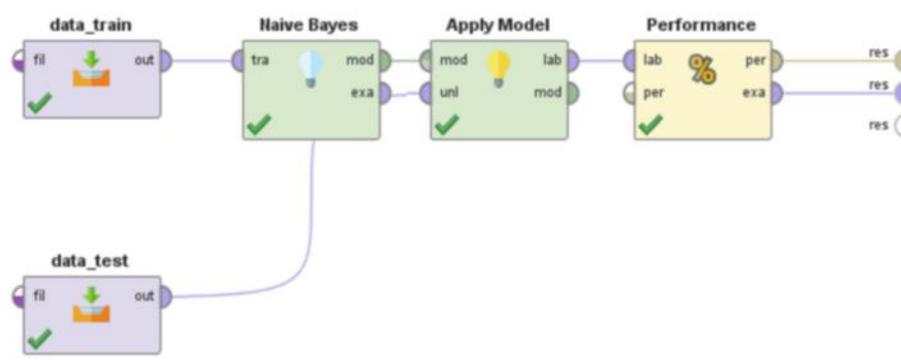
Selanjutnya penulis akan mengganti operator Decision tree dengan operator KNN dan Naïve Bayes untuk mencari nilai Accuracy dari masing-masing Teknik tersebut, lihat Gambar 7 dan 8.



Gambar 6: Proses Mining Menggunakan Teknik Decision tree



Gambar 7: Proses Mining Menggunakan Teknik kNN



Gambar 8: Proses Mining Menggunakan Teknik Naïve Bayes

Hasil Dan Pembahasan

Berikut adalah hasil pengujian data narapidana dengan 3 teknik klasifikasi yang berbeda sebagai berikut :

1. Pengujian dilakukan dengan kebutuhan perangkat hardware dan software sesuai yang tercantum dalam Kebutuhan Perangkat;
2. Data narapidana diambil dari Sistem Database Pemasarakatan dimana penulis telah diberikan ijin oleh Direktorat Jenderal Pemasarakatan yang diwakili oleh Direktur Teknologi Informasi dan Kerja Sama;
3. Data narapidana berasal dari salah satu Lembaga pemasarakatan di Indonesia. Data yang diambil hanya data narapidana dengan kasus narkoba;
4. Data yang diambil berjumlah 1800 record data. Data tersebut dibagi menjadi data training sejumlah 1100 record data dan data testing sejumlah 700 data;
5. Atribut yang dijadikan label adalah atribut kategori narapidana sesuai dengan judul penelitian ini yaitu klasifikasi narapidana kasus narkoba;
6. Penulis menggunakan 3 teknik klasifikasi yaitu Decision tree, KNN, Naïve Bayes. Penulis membandingkan dari 3 teknik klasifikasi tersebut mana yang paling baik untuk proses mining berdasarkan nilai accuracy;

accuracy: 98.86%

	true Resiko Rendah	true Resiko Tinggi	class precision
pred. Resiko Rendah	547	0	100.00%
pred. Resiko Tinggi	8	145	94.77%
class recall	98.56%	100.00%	

Gambar 9: Nilai Accuracy Dengan Teknik Decision tree

accuracy: 92.57%

	true Resiko Rendah	true Resiko Tinggi	class precision
pred. Resiko Rendah	548	45	92.41%
pred. Resiko Tinggi	7	100	93.46%
class recall	98.74%	68.97%	

Gambar 10: Nilai Accuracy Dengan Teknik kNN

accuracy: 68.43%

	true Resiko Rendah	true Resiko Tinggi	class precision
pred. Resiko Rendah	414	80	83.81%
pred. Resiko Tinggi	141	65	31.55%
class recall	74.59%	44.83%	

Gambar 11: Nilai Accuracy Dengan Teknik Naïve Bayes

7. Dari hasil pengujian diatas yang berdasarkan nilai accuracy diperoleh nilai accuracy dari masing-masing Teknik klasifikasi dengan Tabel 7.

Tabel 7: Tabel Hasil Pengujian

NO	TEKNIK KLASIFIKASI	NILA ACCURACY
1	DECISION TREE	98.86%
2	KNN	92.57%
3	NAÏVE BAYES	68.43%

Dari Tabel 7 menunjukkan Teknik *Decision tree* mendapatkan nilai terbesar yaitu 98,86%. Artinya Teknik *Decision tree* paling tepat digunakan untuk proses mining data tersebut.

Tabel 8 adalah sampel prediksi kategori narapidana dengan teknik *Decision tree*. Dari Tabel 8 Muncul prediksi dari kategori narapidana.

Tabel 8: Sample Prediksi Kategori Narapidana (5 Sample)

Row	Kategori Narapidana	Prediction (Kategori Narapidana)	confidence (Resiko Tinggi)	confidence (Resiko Rendah)	Nama	Usia	Kategori Usia	Lama Pidana (tahun)	Jenis Kejahatan	Residivis	Jenis Kejahatan Narkotika
1	Resiko Rendah	Resiko Rendah	0.0	1.0	NAPI 1	45.0	Dewasa	5.0	Narkotika	Tidak	Pengguna
2	Resiko Tinggi	Resiko Tinggi	1.0	0.0	NAPI 2	57.0	Dewasa	7.0	Narkotika	Ya	Pengedar
3	Resiko Rendah	Resiko Rendah	0.0	1.0	NAPI 3	32.0	Dewasa	5.0	Narkotika	Tidak	Pengedar
4	Resiko Rendah	Resiko Rendah	0.0	1.0	NAPI 4	38.0	Dewasa	5.0	Narkotika	Tidak	Pengguna
5	Resiko Rendah	Resiko Rendah	0.016666666666666666	0.9833333333333333	NAPI 5	37.0	Dewasa	5.0	Narkotika	Ya	Pengguna

Penutup

Komparasi teknik *Decision tree*, KNN dan Naïve Bayes dalam klasifikasi narapidana kasus narkotika pada Lembaga pemasyarakatan X telah berhasil dilakukan. Berdasarkan nilai accuracy, teknik *Decision tree* yang paling cocok digunakan untuk proses mining data tersebut. Penulis dapat melakukan prediksi kategori terhadap narapidana kasus narkotika berdasarkan atribut Lama Pidana(tahun), Residivis dan jenis kejahatan narkotika. Mengingat masih banyaknya hal-hal yang belum dapat diimplemetasikan dari penelitian ini, maka beberapa saran yang perlu dipertimbangkan untuk penelitian selanjutnya yaitu :

1. Penelitian ini membandingkan 3 teknik klasifikasi dan hanya berdasarkan nilai accuracy. Penulis menyarankan untuk pengembangan penelitian ini dapat ditambahkan indikator lain seperti nilai precision, recall dan AUC.
2. Dengan jumlah narapidana yang ada pada lapas tentu masih banyak data yang ada pada lapas. Hal ini dikarenakan perekaman data narapidana dari data pribadi, jenis perkara, masa menjalani hukuman, masa pengurangan hukuman, pembinaan, pelanggaran dan pengunjung telah diinput oleh pihak lapas. Hal ini merupakan bagian dari proses bisnis di Pemasyarakatan dimana narapidana yang masuk kedalam lapas kemudian menjalani program pembinaan di dalam lapas hingga bebas. Dalam kebijakan yang ada pada Direktorat Jenderal Pemasyarakatan data di Sistem Database Pemasyarakatan belum pernah dilakukan penghapusan data. Pada penelitian selanjutnya dapat mengambil topik lainnya seperti jenis kejahatan selain narkotika, program pembinaan dan layanan kunjungan narapidana. Terutama untuk layanan

kunjungan narapidana kasus narkotika atau kasus terorisme. Hal ini dapat dijadikan bahan kajian terhadap siapa saja pengunjung yang mengunjungi narapidana kasus narkotika dan terorisme. Sehingga dari aparat penegak hukum yang membidangi tindak pidana narkotika dan terorisme dapat melakukan pelacakan dan prediksi jaringan narkotika dan terorisme.

3. Penelitian ini dilakukan pada lapas yang di salah satu kota besar di Indonesia yang narapidananya beragam suku dan budaya yang ada di Indonesia. Pada penelitian selanjutnya dapat dilakukan lokasi pada lapas yang ada di daerah lainnya yang kebanyakan narapidananya berasal dari suku dan budaya dari daerah tersebut. Hal ini tentu dapat dijadikan bahan kajian terhadap jenis kejahatan apa saja yang banyak terjadi di daerah tersebut. Sehingga dari aparat penegak hukum dapat memberikan pembinaan dan penegakan hukum terhadap penduduk yang ada di daerah tersebut.

Daftar Pustaka

- [1] Artha I Gede, I Wayan Wiryawan, Pengendalian Peredaran Gelap Narkotika Oleh Narapidana Dari Dalam Lembaga Pemasyarakatan (LAPAS), Jurnal Magister Hukum Udayana Vol. 4 No. 3 : 588 - 602, September 2015
- [2] Pemerintah Indonesia.2012, PP Nomor 99 Tahun 2012 tentang Perubahan Kedua Atas Peraturan Pemerintah Nomor 32 Tahun 1999 tentang Syarat dan Tata Cara Pelaksanaan Hak Warga Binaan Pemasyarakatan (Konsideran Menimbang huruf a), Presiden Republik

Indonesia : H Susilo Bambang Yudhoyono, Jakarta:

- [3] Novarizal Riky, Narapidana Narkotika Di Lembaga Pemasyarakatan Umum Studi Terhadap Tiga Narapidana Kategori Bandar dan Pengedar di Lapas Kelas IIA Pekanbaru, Jurnal Sisi Lain Realita:Vol. 1 No. 1 (2016)
- [4] Haryono, Kebijakan Perlakuan Khusus Terhadap Narapidana Risiko Tinggi Di Lembaga Pemasyarakatan (Studi Kasus Di Lembaga Pemasyarakatan KLS III GN. Sindur), JIKH Vol. 11 No. 3 November 2017 : 231 – 247
- [5] Davies, and Paul Beynon, 2004, "Database Systems Third Edition", Palgrave Macmillan, New York.
- [6] Han, J., dan Kamber, M. 2006. "Data mining Concepts and Techniques Second Edition". San Francisco: Morgan Kauffman.
- [7] Wulandari, Retno Tri. 2017. "Data mining Teori dan Aplikasi Rapid-Miner".Yogyakarta:Penerbit Gava Media.
- [8] Olson & Delen. Advanced Data mining Techniques. USA: Springer-Verlag Berlin Heidelberg. 2008. [9] Wu. (2009). "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemo-tor"
- [9] Ashleigh LaCourse, Shelley Johnson Listwan, Shannon Reid, and Jennifer L. Hartman, "Recidivism and Reentry: The Role of Individual Coping Styles", Crime & Delinquency, Vol. 65(1) 46– 68,2019.
- [10] Daniel P. Mears, Joshua C. Cochran, William D. Bales, Avinash S. Bhati, "Recidivism and Time Served in Prison", The Journal Of Criminal Law & Criminology Vol. 106, No. 1, 2016.