

PENERAPAN KONSEP BUSINESS INTELLIGENCE UNTUK PERCEPATAN PENYELESAIAN PERKARA PADA PANMUD PERDATA KHUSUS MAHKAMAH AGUNG RI

Astria Yumalia¹, Richardus Eko Indrajit²

Mahkamah Agung RI¹, ABFI Institute Perbanas²
Jakarta, Indonesia

E-mail : astria.yumalia@mahkamahagung.go.id¹, indrajit@post.harvard.edu²

ABSTRAK

Surat Keputusan Ketua MA No. 214/KMA/SK/XII/2014 tentang Jangka Waktu Penanganan Perkara di Mahkamah Agung yang kini tengah menghadapi lonjakan jumlah perkara. Kelambatan penyelesaian perkara meningkatkan tumpukan perkara dan penyelesaian perkara yang lama berimplikasi dengan semakin besar biaya disertai dengan prosedur panjang menimbulkan kerugian dan ketidakpastian hukum bagi pencari keadilan. *Business intelligence* merupakan sistem aplikasi yang mampu menganalisa data operasional dan data transaksional di masa lampau ke dalam bentuk *knowledge* untuk mendukung keputusan dan perencanaan organisasi. *Business Intelligence* sebagai alat bantu untuk mengolah dan menganalisa data register perkara perdata khusus sehingga dapat memberikan informasi yang akurat. Penelitian dilakukan dengan menerapkan konsep *business intelligence* dengan pengklasifikasian algoritma yang tepat dalam percepatan penyelesaian perkara. Beberapa algoritma yang diuji, yaitu algoritma *decission tree*, *Naive Bayes*, *Logistic Regression*, *Decision Stump* dan *k-NN*. Validasi dilakukan dengan menggunakan *10 fold cross validation*, *Area Under Curve (AUC)* untuk model evaluasi kinerja. Metode klasifikasi yang diusulkan dilihat dari perbandingan akurasi dan AUC, lalu melakukan uji t-test untuk analisis uji beda. Hasil yang diperoleh menunjukkan persentase perkara yang diselesaikan tepat waktu.

Kata kunci : *Busniess intelligence*, algoritma klasifikasi, dan komparasi.

1. PENDAHULUAN

Surat Keputusan Ketua MA No. 214/KMA/SK/XII/2014 tentang Jangka Waktu Penanganan Perkara di Mahkamah Agung (Agung & Indonesia, 2010) yang kini tengah menghadapi lonjakan jumlah perkara (Nurhadi, 2015)(Agung, 2014). Kelambatan penyelesaian perkara meningkatkan tumpukan perkara dan penyelesaian perkara yang lama berimplikasi dengan semakin besar biaya disertai dengan prosedur panjang menimbulkan kerugian dan ketidakpastian hukum bagi pencari keadilan. Kecepatan waktu memeriksa perkara akan berkorelasi dengan meningkatnya produktifitas dalam memutus perkara, sedangkan tingginya produktifitas akan berkorelasi dengan berkurangnya sisa perkara.

Percepatan penyelesaian perkara dari waktu ke waktu menghasilkan data dan informasi dalam bentuk laporan yang perlu di analisis oleh panmud perdata khusus dalam rangka mendukung keputusan. Namun, itu adalah tugas yang sulit bagi manusia untuk menganalisis data yang kompleks (Turban, Sharda, Delen, & King, 2011). Kesulitan ini menyebabkan perkembangan teknik *business intelligence*, yang bertujuan mengekstraksi

pengetahuan yang berguna untuk mendukung pengambilan keputusan.

Dalam percepatan penyelesaian perkara, menggunakan data register perkara perdata khusus yang tersedia saat ini, dapat memperkirakan jumlah perkara yang harus diselesaikan tepat waktu, dan penggunaan informasi tersebut dalam membuat keputusan panmud perdata khusus.

Selain itu, panmud perdata khusus memiliki sedikit pengetahuan tentang *data mining*, maka ada kebutuhan untuk mengembangkan suatu kerangka kerja yang disederhanakan untuk membantu panmud perdata khusus dalam memanfaatkan metode *data mining* untuk percepatan penyelesaian perkara.

Salah satu cara yang efektif untuk menganalisa laporan dari kampanye sebelumnya dan serupa dalam mencari tren dan pola adalah melalui *business intelligence* dan teknik *data mining*, untuk membangun model dan kemudian ekstrak pengetahuan (Witten, Frank, & Hall, 2011). *Business intelligence* adalah sebuah konsep luas yang mencakup data mining yang terdiri dalam ekstraksi pengetahuan dari data mentah.

Business intelligence merupakan sistem aplikasi yang mampu menganalisa data operasional dan data transaksional di masa lampau ke dalam bentuk *knowledge* untuk mendukung keputusan dan

perencanaan organisasi. *Business Intelligence* sebagai alat bantu untuk mengolah dan menganalisa data register perkara perdata khusus sehingga dapat memberikan informasi yang akurat.

Penelitian dilakukan dengan menerapkan konsep *business intelligence* dengan pengklasifikasian algoritma yang tepat dalam percepatan penyelesaian perkara.

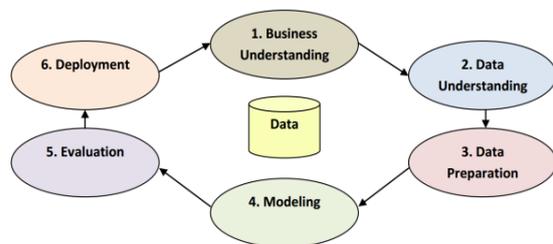
Dalam penulisan ini dibagi menjadi beberapa bagian. Pada bagian 2, menjelaskan tentang landasan teori. Pada bagian 3, menjelaskan metode yang diusulkan. Hasil penelitian dan pembahasan mengenai komparasi metode yang diusulkan dijelaskan dalam bagian 4. Penutup, pada bagian ini akan menjelaskan tentang kesimpulan dan saran dari penelitian.

2. METODOLOGI

Penelitian ini bertujuan untuk membandingkan performa beberapa metode klasifikasi untuk mengetahui tingkat akurasi dalam percepatan penyelesaian perkara dengan menerapkan konsep business intelligence.

Desain Penelitian

Metode yang diusulkan pada paper ini adalah CRISP-DM. The Cross-Industry Standard Process for Data Mining (CRISP-DM) was developed in 1996 (Chapman, 2000). Terdapat 6 langkah tahapan CRISP-DM, yaitu: 1) Business Understanding, 2) Data Understanding, 3) Data Preparation, 4) Modeling, 5) Evaluation, 6) Deployment.



Gambar 2.1. Tahapan CRISP-DM

Business Understanding

Tahap ini adalah pemahaman tentang masalah secara umum baik dari bisnis atau untuk penelitian apa yang akan dilakukan. Serta batasan-batasan masalah yang akan di selesaikan dengan data mining.

Data Understanding

Merupakan tahap mengumpulkan data, memahami data yang didapat lalu mengecek data tersebut apakah sudah sesuai atau belum untuk nantinya digunakan.

Data Preparations

Melakukan persiapan set data akhir, yang akan digunakan untuk fase berikutnya dengan memilih dan mengeleminasi atribut apa saja yang akan digunakan dan tidak untuk data mining.

Modeling

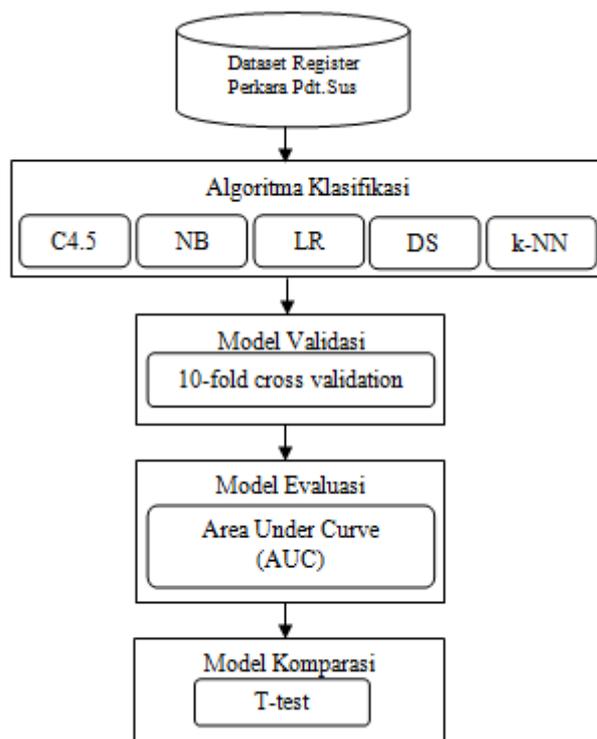
Memilih dan menerapkan teknik pemodelan data mining yang sesuai dengan permasalahan yang akan di selesaikan sehingga dapat diperoleh hasil yang optimal.

Evaluation

Melakukan validasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yang terdapat dalam rapidminer yang disesuaikan dengan tujuan penelitian. Selanjutnya melakukan pengambilan keputusan mengenai penggunaan hasil data mining.

Deployment

Menentukan hasil penelitian yang sudah dilakukan baik dalam bentuk laporan atau rekomendasi untuk instansi terkait sesuai tujuan penelitian. Atau untuk rekomendasi pengambilan keputusan.



Gambar 2.2. Kerangka Kerja yang diusulkan

Data Set

Dataset yang digunakan dalam pengujian ini adalah data register kasasi perkara perdata khusus pada tahun 2015 yang digunakan untuk mengukur tingkat akurasi masing-masing algoritma. Jumlah record sebanyak 854 perkara.

Tabel 2.1. Dataset Perkara Perdata Khusus Mahkamah Agung RI

| No. Perkara | No. Register | Identifikasi | Partai | TP | FP | FN | TN | Accuracy | Precision | Recall | F1 Score | Area Under Curve |
|-------------|--------------|--------------|--------|----|----|----|----|----------|-----------|--------|----------|------------------|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |

Algoritma Klasifikasi

Penelitian ini akan membandingkan beberapa algoritma yang telah dikenal secara umum, antara lain decision tree, Naive Bayes, K-NN, Random Forest dan Decision Stump, yang kemudian akan didapatkan akurasi terbaik untuk setiap performance.

Model Validasi

Pada Penelitian ini dilakukan pengujian dengan 10-fold cross-validation. Hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa 10-fold crossvalidation adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat. 10-fold cross validation akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 pengujian. Proses pengujian metode dimulai dari pembagian dataset dengan metode 10-fold cross validation yang membagi dataset menjadi dua yaitu data training dan data testing.

Tabel 2.2. Stratified 10 Fold Cross Validation (Wahono, Herman, & Ahmad, 2014)

| n-validation | Dataset's Partition | | | | | | | | | |
|--------------|---------------------|---|---|---|---|---|---|---|---|---|
| 1 | █ | | | | | | | | | |
| 2 | | █ | | | | | | | | |
| 3 | | | █ | | | | | | | |
| 4 | | | | █ | | | | | | |
| 5 | | | | | █ | | | | | |
| 6 | | | | | | █ | | | | |
| 7 | | | | | | | █ | | | |
| 8 | | | | | | | | █ | | |
| 9 | | | | | | | | | █ | |
| 10 | | | | | | | | | | █ |

Model Evaluasi

Makalah ini menggunakan tabel Akurasi dan Area Under Curve (AUC) sebagai indikator pada pengujian akurasi model klasifikasi yang digunakan untuk mengevaluasi performa dari model tersebut (Lessman dkk, 2008). Pada tabel confusion matrix berisi nilai false positive (FP), false negative (FN), true positive (TP), true negative (TN). Pengujian dan AUC biasa digunakan untuk mengukur akurasi dari metode data mining klasifikasi. Berikut ini merupakan kriteria dalam pengujian AUC.

Selanjutnya diterapkan tahapan evaluasi menggunakan Area Under Curve (AUC) untuk mengukur hasil akurasi dari performa model

klasifikasi. Hasil akurasi dilihat menggunakan curva Receiver Operating Characteristic (ROC) dan hasil confusion matrix. ROC menghasilkan dua garis dengan bentuk true positive sebagai garis vertikal dan false positive sebagai garis horizontal. Pengukuran akurasi dengan confusion matrix dapat dilihat pada tabel berikut ini:

Tabel 2.3. Confusion Matrix

| | | Aktual | |
|------------|--------|---------------------|--------------------|
| | | True | False |
| Predi cted | T rue | True Positive (TP) | False Negative(FN) |
| | F alse | False Negative (FN) | True Negative (TN) |

Formulasi perhitungan adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$Sensitivity = TP_{rate} = \frac{TP}{TP + FN}$$

$$Spesificity = TN_{rate} = \frac{TN}{TN + FP}$$

$$FP_{rate} = \frac{FP}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F Measure = \frac{2RP}{R + P}$$

$$G - Mean = \sqrt{sensitivity * specificity}$$

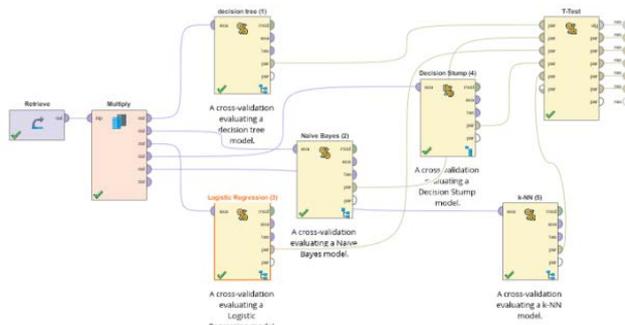
Evaluasi dalam penelitian ini adalah menggunakan uji t (t-test). Uji t adalah membandingkan hubungan antara dua variabel yaitu variabel respon dan variabel predictor. Uji t sample berpasangan (paired-sample t-test) digunakan untuk menguji perbandingan selisih dua rata-rata dari sua sample yang berpasangan dengan asumsi bahwa data terdistribusi normal.

Tabel 2.4. Nilai AUC dan Keterangan

| Nilai AUC | Klasifikasi |
|-------------|----------------|
| 0.90 - 1.00 | Paling Baik |
| 0.80 - 0.90 | Baik |
| 0.70 - 0.80 | Adil atau Sama |
| 0.60 - 0.70 | Rendah |
| 0.50 - 0.60 | Gagal |

Model Perbandingan

Model komparasi dari makalah ini yaitu dengan menggunakan metode parametrik. Metode parametrik itu sendiri menggunakan pengujian t-test untuk mendapatkan model terbaik dari pengujian yang dilakukan terhadap beberapa model klasifikasi tersebut.



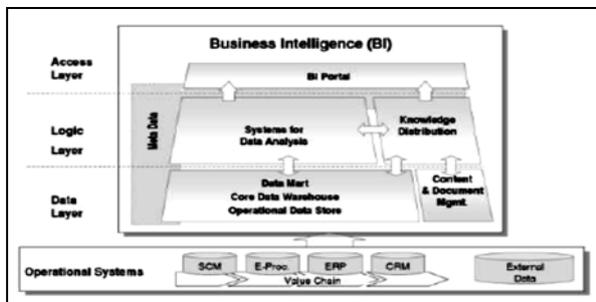
Gambar 3.3. Uji Beda dengan t-Test Menggunakan RapidMiner

3. LANDASAN TEORI

Business Intelligence

Business Intelligence merupakan kerangka kerja konseptual untuk mendukung keputusan bisnis, *business intelligence* menggabungkan arsitektur, basisdata atau *data warehouse*, tool analisis dan aplikasi. *Business Intelligence* digunakan untuk aplikasi dan teknologi dalam mengumpulkan, menyimpan, menganalisa, dan menyediakan akses pada data sehingga dapat membantu pengguna dari kalangan perusahaan atau organisasi untuk mengambil keputusan dengan lebih baik dan tepat.

Business Intelligence dalam kaitannya *management support* terhadap data terstruktur dan data tidak terstruktur, merupakan proses mengintegrasikan dan menyatukan komponen-komponen untuk menangani data-data pada *business intelligence framework*. Pendekatan tersebut akan dilakukan dengan tiga jenis pendekatan yaitu mengintegrasikan data yang terstruktur dan tidak terstruktur, melakukan analisis koleksi data dan melakukan pendistribusikan hasil analisis ke dalam bentuk yang sesuai dengan kebutuhan. Pendekatan tersebut diatas dapat memanfaatkan tiga lapisan *business intelligence framework* berupa *data layer*, *logic layer* dan *access layer* seperti yang ditunjukkan pada gambar 3.2 arsitektur *business intelligence* dengan beberapa lapisan.

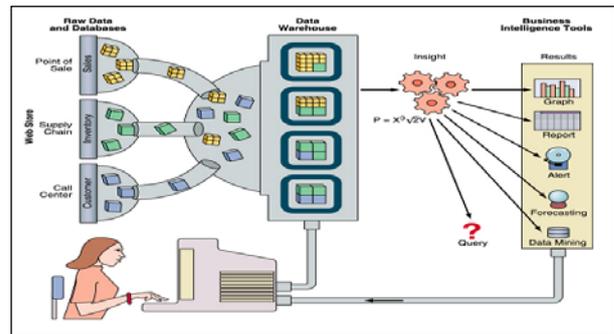


Gambar 3.2 Arsitektur Business Intelligence

Dalam sebuah perusahaan atau organisasi, keputusan dibuat pada tingkatan atau level yang berbeda-beda, keputusan juga harus dilakukan secepat mungkin untuk mempertahankan daya saing,

sehingga dapat membuat keputusan yang benar dan menjadi dasar yang solid dalam hal data, informasi dan ketersediaan pengetahuan. Dasar ini dapat bersumber dari informasi atau data berita bisnis, konferensi, pelanggan, tenaga penjualan, dan sebagainya. Sehingga bersama-sama membentuk sebuah konsep mengenai pasar dan dunia bisnis tertentu.

Gambar 3.3 menyajikan pemahaman dasar mengenai sistem *business intelligence*. Sebuah sistem *business intelligence* dengan kata lain merupakan kombinasi *data warehouse* dan sistem pendukung keputusan. Hal ini menjelaskan bagaimana data dari sumber-sumber yang berbeda dapat diekstraksi dan disimpan dan selanjutnya diambil untuk dianalisis. Kegiatan utama *business intelligence* meliputi pengumpulan, menyiapkan dan menganalisa data. Dalam proses *business intelligence* data yang digunakan harus berkualitas tinggi, dengan cara memperolehnya dari berbagai sumber data yang dikumpulkan, kemudian diubah, lalu dibersihkan, selanjutnya dimuat dan disimpan dalam basisdata *data warehouse*.



Gambar 3.3 Pemahaman Dasar Sistem Business Intelligence (Ranjan, 2009)

Arsitektur dari sebuah sistem *business intelligence* terdiri atas enam komponen utama (Vercellis, 2009) yaitu :

1. Data Source

Pada tahap pertama ini diperlukan proses untuk mengumpulkan dan mengintegrasikan data yang disimpan dalam berbagai sumber yang bervariasi yang saling berbeda baik itu asal maupun jenisnya. Sumber data ini berasal dari data yang terdapat pada *operational system*, tetapi juga bisa berasal dari dokumen yang tidak terstruktur seperti email dan data yang dikirimkan oleh pihak luar.

2. Data Warehouse

Pada tahap ini proses menggunakan *extraction* dan *transformation tool* yang dikenal sebagai ETL (*Extract, Transform, Load*), data yang berasal dari berbagai sumber yang berbeda disimpan ke dalam basisdata yang ditujukan untuk mendukung proses analisis *business intelligence*.

3. *Data Exploration*
Pada tahap ini, *tools* yang berfungsi untuk keperluan analisis *business intelligence* pasif digunakan. *Tools* ini terdiri dari *query* dan *reporting system*, serta *statistical methods*. Metodologi ini bersifat pasif dikarenakan para pengambil keputusan harus mengambil keputusan berdasarkan hipotesis mereka sendiri atau mendefinisikan kriteria dari *data extraction*, kemudian menggunakan *tools* analisis untuk menemukan jawaban dan mencocokkannya dengan hipotesa awal mereka.
4. *Data Mining*
Pada tahap ini proses terdiri sejumlah metodologi *business intelligence* bersifat aktif yang tujuannya untuk mengekstrak informasi dan pengetahuan dari data tersebut. Metodologi ini berisi sejumlah model matematika untuk pengenalan pola (*pattern*), pembelajaran mesin (*machine learn*) dan teknik *data mining*.
5. *Optimization*
Pada tahap ini menghasilkan solusi dimana solusi terbaik harus dipilih dari sekian solusi alternatif yang ada, dan biasanya sangat banyak dan beragam atau bervariasi.
6. *Decisions*
Pada tahap ini yang menjadi persoalan utama merupakan bagaimana menentukan keputusan akhir yang akan diambil yang dikenal sebagai *decision making process*. Walaupun metodologi *business intelligence* berhasil diterapkan, pilihan untuk mengambil sebuah keputusan tetap ada ditangan para pengambil keputusan tersebut.

Menurut (Turban et al., 2011), *business intelligence* terbagi ke dalam lima jenis atau kategori yaitu :

1. *Enterprise Reporting* yakni digunakan untuk menghasilkan laporan-laporan statis yang didistribusikan ke banyak orang. Jenis laporan ini sangat sesuai untuk laporan operasional dan *dashboard*.
2. *Cube Analysis* yakni digunakan untuk menyediakan analisis OLTP multidimensional yang ditujukan untuk manajer bisnis dalam lingkungan terbatas.
3. *Ad Hoc Query and Analysis* yakni digunakan untuk memberikan akses kepada *user* agar dapat melakukan *query* pada basis data, dan menggali informasi sampai pada tingkat paling dasar dari informasi transaksional. *Query* ini berfungsi untuk mengeksplor informasi yang dilakukan oleh *user*.
4. *Statistical Analysis and Data Mining* yakni digunakan untuk melakukan analisis prediksi atau menentukan korelasi sebab akibat diantara dua matrik.

5. *Delivery Report and Alert* yakni digunakan secara proaktif untuk mengirimkan laporan secara lengkap atau memberikan peringatan kepada populasi *user* yang besar atau banyak.

Data Mining

Data Mining merupakan suatu metode penambangan data atau penemuan data dan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data Mining dikenal pula sebagai *knowledge discovery in database* atau disingkat KDD yakni berupa kegiatan yang meliputi pengumpulan, pemakaian data, historis data untuk mencari dan menemukan keteraturan, pola atau hubungan dalam suatu set data berukuran besar. Data Mining dapat pula disebut sebagai serangkaian proses atau tahapan untuk menggali dan mencari nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data. Data Mining mempunyai keterkaitan dengan bidang keilmuan lainnya seperti sistem basisdata, data warehouse, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining juga didukung oleh keilmuan lain seperti neural network, pengenalan pola, spatial data analysis, image database, dan signal processing. Data Mining merupakan kegiatan atau proses menemukan polapola dalam data dan dalam menemukan pola yang menarik tersebut bersumber dari data dalam jumlah besar, data tersebut tersimpan dalam basisdata, data warehouse, atau teknologi penyimpanan informasi lainnya.

Serangkaian proses untuk mendapatkan pengetahuan atau pola dari sekumpulan data disebut dengan data mining (Witten et al., 2011). Klasifikasi merupakan topik penelitian yang penting dalam data mining, karena teknik klasifikasi dapat memecahkan masalah dengan menganalisis data dalam database yang masing-masing data dikelompokkan kedalam kelas tertentu.

Secara umum, klasifikasi adalah fungsi data mining yang menggambarkan dan membedakan kelas data atau konsep. Tujuan dari klasifikasi adalah untuk secara akurat memprediksi label kelas dari instance yang nilai atributnya diketahui, tapi nilai kelasnya tidak diketahui.

Performa algoritma data mining dalam banyak kasus tergantung pada kualitas dataset, karena data training berkualitas rendah dapat menyebabkan klasifikasi yang lemah (Endres & Rombach, 2003). Dengan demikian, dibutuhkan teknik data preprocessing untuk mempersiapkan data yang akan diproses. Hal ini dapat meningkatkan kualitas data, sehingga membantu untuk meningkatkan akurasi dan efisiensi proses data mining. Beberapa teknik data preprocessing diantaranya adalah data cleaning: menghapus data yang mengandung error, data integration: menggabungkan data dari berbagai sumber, data transformation: normalisasi data, dan data reduction: mengurangi ukuran data dengan

menggabungkan dan menghilangkan fitur yang berlebihan.

Decision Tree (C4.5)

Decision Tree adalah algoritma klasifikasi yang dinyatakan sebagai partisi rekursif dari ruang contoh. Decision Tree terdiri dari node yang membentuk pohon berakar, yang berarti pohon diarahkan dengan simpul yang disebut akar. Sebuah node dengan tepi keluar disebut internal atau tes node. Semua node yang lain disebut daun. Dalam pohon keputusan, setiap simpul internal membagi ruang misalnya menjadi dua atau lebih sub ruang sesuai dengan fungsi diskrit tertentu dari atribut nilai.

Sebuah pohon keputusan terdiri dari internal node yang menentukan tes pada variabel masukan individu atau atribut yang membagi data menjadi himpunan bagian yang lebih kecil, dan serangkaian node daun menetapkan kelas untuk masing-masing pengamatan di segmen yang dihasilkan. Pada penelitian ini, C4.5 membangun pohon keputusan dengan menggunakan konsep entropi informasi. Entropi sampel S dari pengamatan yang diklasifikasikan diberikan oleh:

$$Entropy(S) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$$

Dimana p_1 adalah proporsi dari kelas yang nilainya 1 dan p_0 proporsi dari kelas yang nilainya 0 dalam sample S . C4.5 memeriksa informasi *Gain* normalisasi (perbedaan entropi) yang dihasilkan dari memilih atribut untuk membagi data. Atribut dengan informasi *Gain* normalisasi tertinggi adalah yang digunakan untuk membuat keputusan. Algoritma kemudian terbagi lagi menjadi subset lebih kecil (Brown & Mues, 2012).

Naïve Bayes

Naïve Bayes didasarkan pada teorema Bayes. Oleh karena itu, meninjau teorema Bayes dan kemudian menggambarkan klasifikasi. List dari paket *software* data mining yang mendukung pembelajaran klasifikasi Naïve Bayes tersedia. Beberapa aplikasi klasifikasi Naïve Bayes tersedia dengan referensi (Fu, 1997). Teorema Bayes berasal dari persamaan:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Klasifikasi Naïve Bayes memperkirakan persamaan probabilitas berikut:

$$P(y) = \frac{n_y}{n}$$

$$P(x_i|y) = \frac{n_{y \& x_i}}{n_y}$$

Keterangan:

n , total nomor dari point data pada data set training
 n_y , nomor dari point data target class y

$n_{y \& x_i}$ nomor dari point data dengan target class y
 i , variabel atribut yang mengambil nilai dari x_i

3.5. Logistic Regression

Logistik regresi (LOG) memperluas ide-ide dari regresi linier berganda dengan situasi di mana variabel dependen, y , diskrit. Dalam regresi logistik tidak ada asumsi yang dibuat mengenai distribusi variabel independen. Prosedur kemungkinan maksimum dapat diadopsi untuk mendapatkan parameter. Mengingat sampel N dan label kelas diamati mereka, fungsi log-likelihood ditampilkan di bawah dan dapat dimaksimalkan dengan menggunakan algoritma Newton-Raphson (Yu, Chen, Koronios, Zhu, & Guo, n.d.).

Decision Stump

Decision Stump pada dasarnya merupakan pohon keputusan dengan lapisan tunggal. Perbandingan dari pohon yang memiliki beberapa lapisan, Decision Stump pada dasarnya berhenti setelah split pertama. Decision Stump biasanya digunakan dalam segmentasi populasi untuk data yang besar. Kadang-kadang, juga digunakan untuk membantu membuat simplexes atau tidak ada model keputusan untuk data yang lebih kecil dengan sedikit data. Decision Stump umumnya lebih mudah untuk dibangun dibandingkan dengan Decision Tree. Pada saat yang sama, coding SAS untuk Decision Stump lebih mudah dikelola dibandingkan dengan CART dan CHAID. Alasannya adalah bahwa Decision Stump adalah hanya satu single run dari algoritma tree dan dengan demikian tidak perlu mempersiapkan data untuk split berikutnya. Pada saat yang sama, tidak butuh menentukan data untuk split berikutnya yang membuat penggantian nama dari pengelolaan output sederhana (Undavia, 2014).

K Nearest Neighbour (k-NN)

K-nearest neighbour merupakan metode klasifikasi pertama yang menempatkan point k data yang paling mirip dengan titik data k terdekat untuk menentukan kelas target titik terdekat. Untuk menentukan k-nearest neighbour dari titik data perlu mengukur persamaan atau perbedaan antara titik data, titik data tersebut dapat diukur dengan beberapa teknik misalnya, Euclidean distance, Minkowski distance, Hamming distance, koefisien korelasi Pearson, dan persamaan cosine (Fu, 1997). Euclidean distance didefinisikan sebagai berikut:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^p (x_{i,l} - x_{j,l})^2}, i \neq j$$

Euclidean distance adalah teknik membandingkan perbedaan dari dua data (x_i dan x_j). Semakin besar jarak Euclidean, semakin beda dua

point data, dan dua titik data yang jauh terpisah dipisahkan dalam ruang data dimensi p .

Minkowski didefinisikan sebagai berikut:

$$d(x_i, x_j) = \left(\sum_{l=1}^p |x_{i,l} - x_{j,l}|^r \right)^{1/r}, i \neq j$$

Berikut ini merupakan koefisien korelasi Pearson p :

$$\rho_{x_i x_j} = \frac{Sx_i x_j}{Sx_i Sx_j}$$

Persamaan cosine menganggap dua titik data x_i dan x_j sebagai dua vektor dalam ruang dimensi p dan menggunakan cosinus dari sudut θ antara dua vektor untuk mengukur kesamaan dari dua titik data:

$$\cos(\theta) = \frac{x_i x_j}{\|x_i\| \|x_j\|}$$

4. HASIL DAN PEMBAHASAN

Eksperimen yang dilakukan dalam penelitian ini menggunakan sebuah platform komputer berbasis Intel(R) Core(TM) i3-2330M CPU @ 2.20GHz/1.50GHz, RAM 4,00 GB, dan sistem operasi Microsoft Windows 10 Home 64-bit. Sedangkan lingkungan pengembangan aplikasi menggunakan tools rapid miner 7.0 library.

Business Understanding

Pengolahan data dilakukan untuk mengolah dan menganalisa data register perkara perdata khusus sehingga dapat memberikan informasi yang akurat,

Data Understanding

Dalam penelitian ini menggunakan data data register perkara kasasi perdata khusus pada tahun 2015.

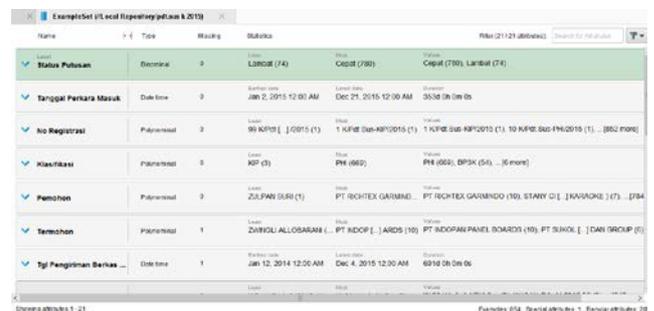
Tabel 4.1. Atribut dan Data Perkara Perdata Khusus Mahkamah Agung RI

| Atribut | Type Atribut | Keterangan |
|---------------------------------------|--------------|--|
| Tanggal Perkara Masuk | Date Time | Tanggal Perkara diregister |
| No Registrasi | Polynomial | Nomor registrasi perkara para pemohon/penggugat dan termohon/tergugat |
| Klasifikasi | Polynomial | Klasifikasi perkara yang terdiri dari Arbitrase, BPSK, HKI, KIP, KPPU, Pailit, Parpol, dan PHI |
| Pemohon | Polynomial | Nama Pemohon atau sebagai Penggugat |
| Termohon | Polynomial | Nama Termohon atau sebagai Tergugat |
| Tgl Pengiriman Berkas Dari PN | Date Time | Tanggal pengiriman berkas dari pengadilan pengaju ke tingkat kasasi |
| No Surat Pengantar Pengadilan Pengaju | Polynomial | Nomor surat pengantar dari pengadilan pengaju |
| Asal Pengadilan Pengaju | Polynomial | Berkas yang berasal dari pengadilan pengaju |
| No Putusan Tingkat 1 | Polynomial | Nomor registrasi perkara di tingkat pertama |

| | | |
|-------------------------------------|------------|---|
| Tanggal Putusan Tingkat 1 | Date Time | Hasil Putusan perkara di tingkat pertama |
| Tanggal Perkara Distribusi | Date Time | Tanggal Perkara didistribusikan ke Majelis untuk disidangkan |
| Amar Putusan | Polynomial | Hasil putusan dari musyawarah ucapan majelis hakim, amar terdiri dari kabul, tolak, N.O, dan pencabutan |
| Tanggal Putusan Kasasi | Date Time | Tanggal diputusnya perkara pada hasil musyawarah ucapan di ruang sidang majelis hakim |
| Nama P1 | Polynomial | Nama pembaca satu |
| Nama P2 | Polynomial | Nama pembaca kedua |
| Nama P3 | Polynomial | Nama pembaca ketiga |
| Nama Panitera Pengganti | Polynomial | Nama panitera pengganti pada askor |
| Tgl Perkara Kembali | Date Time | Tanggal Minutasi yang menyatakan berkas perkara selesai disidang oleh Majelis dan dikembalikan untuk proses pengiriman ke PN yang diadministrasikan di Perdata Khusus |
| Tanggal Kirim Ke Pengadilan Pengaju | Date Time | Tanggal dikirimnya berkas perkara ke pengadilan pengaju |
| Biaya | Integer | Biaya proses berperkara yang harus dikeluarkan oleh para penggugat. Apabila nilai gugatan dibawah Rp. 150 Juta maka biaya perkara dibebankan Negara atau NIHIL |
| Status Putusan | Binominal | Label yang terdiri dari cepat atau lambatnya perkara |

Data Preparations

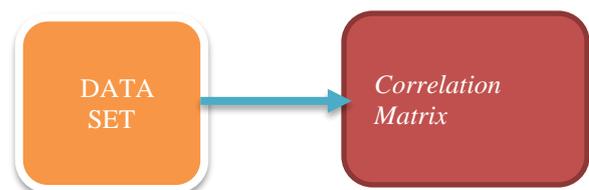
Pada tahapan ini dilakukan proses cleaning untuk menghilangkan missing dan noisy pada dataset.



Gambar 4.1. Cek data noise data Preparation

Modeling

Model Analisa faktor yaitu menganalisis faktor yang berpengaruh dalam mengklasifikasi dan mengetahui status putusan perkara perdata khusus dengan menggunakan *Correlation Matrix* dengan model sebagai berikut:



Gambar 4.2. Model Analisis Faktor

Evaluation

Analisa Faktor Dilihat dari *Correlation Matrix* digunakan untuk mengetahui korelasi antar atribut yang ada pada data register perkara perdata khusus.

| Atribut... | Tanggal... | No Reg... | Klasifikasi... | Pemohon... | Terhadap... | Tgl Pen... | No Surat... | Asal Pe... | No Putu... | Tanggal... | Tanggal... | Actar P... | Tanggal... | Matra... | Matra... |
|----------------|------------|-----------|----------------|------------|-------------|------------|-------------|------------|------------|------------|------------|------------|------------|----------|----------|
| Tanggal... | 1 | 0.906 | 0.919 | 0.871 | 0.864 | 0.955 | 0.990 | 0.179 | 0.905 | 0.966 | 0.988 | 0.148 | 0.961 | 0.158 | 0.873 |
| No Reg... | 0.906 | 1 | 0.917 | 0.874 | 0.868 | 0.952 | 0.986 | 0.174 | 0.903 | 0.960 | 0.983 | 0.155 | 0.958 | 0.160 | 0.873 |
| Klasifikasi... | 0.919 | 0.917 | 1 | 0.924 | 0.832 | 0.932 | 0.919 | -0.085 | 0.917 | 0.965 | 0.914 | -0.006 | 0.943 | 0.227 | -0.229 |
| Pemohon... | 0.871 | 0.874 | 0.924 | 1 | 0.958 | 0.920 | 0.987 | 0.155 | 0.918 | 0.942 | 0.958 | 0.154 | 0.930 | 0.243 | 0.874 |
| Terhadap... | 0.864 | 0.868 | 0.868 | 0.958 | 1 | 0.942 | 0.981 | 0.153 | 0.903 | 0.957 | 0.901 | 0.151 | 0.940 | 0.265 | 0.869 |
| Tgl Pen... | 0.955 | 0.952 | 0.932 | 0.920 | 0.942 | 1 | 0.940 | 0.157 | 0.952 | 0.999 | 0.954 | 0.141 | 0.915 | 0.179 | 0.870 |
| No Surat... | 0.990 | 0.986 | 0.919 | 0.987 | 0.991 | 0.948 | 1 | 0.157 | 0.904 | 0.968 | 0.987 | 0.167 | 0.964 | 0.083 | 0.870 |
| Asal Pe... | 0.179 | 0.174 | -0.085 | 0.155 | 0.153 | 0.157 | 0.157 | 1 | 0.158 | 0.980 | 0.157 | -0.040 | 0.911 | 0.162 | -0.011 |
| No Putu... | 0.905 | 0.903 | 0.917 | 0.874 | 0.868 | 0.952 | 0.986 | 0.174 | 1 | 0.962 | 0.983 | 0.156 | 0.960 | 0.160 | 0.873 |
| Tanggal... | 0.966 | 0.960 | 0.961 | 0.943 | 0.947 | 0.968 | 0.958 | 0.098 | 0.963 | 1 | 0.963 | 0.157 | 0.931 | 0.287 | 0.937 |
| Tanggal... | 0.988 | 0.983 | 0.914 | 0.988 | 0.981 | 0.974 | 0.987 | 0.177 | 0.969 | 0.983 | 1 | 0.148 | 0.981 | 0.158 | 0.876 |
| Actar P... | 0.148 | 0.155 | -0.001 | 0.154 | 0.151 | 0.141 | 0.157 | -0.043 | 0.155 | 0.167 | 0.148 | 1 | 0.148 | 0.159 | 0.017 |
| Tanggal... | 0.961 | 0.959 | 0.943 | 0.936 | 0.948 | 0.911 | 0.954 | 0.151 | 0.919 | 0.921 | 0.967 | 0.148 | 1 | 0.158 | 0.848 |
| Matra... | 0.158 | 0.160 | 0.227 | 0.243 | 0.240 | 0.275 | 0.262 | 0.152 | 0.163 | 0.167 | 0.229 | 0.229 | 0.158 | 1 | -0.016 |
| Matra... | 0.158 | 0.160 | 0.227 | 0.243 | 0.240 | 0.275 | 0.262 | 0.152 | 0.163 | 0.167 | 0.229 | 0.229 | 0.158 | -0.016 | 1 |

Tabel 4.2. *Correlation Matrix*

Deployment

1. Hubungan Antar Faktor Menambah dan menghilangkan atribut dalam dataset seperti atribut yang tidak berpengaruh.
2. Perbandingan Algoritma Klasifikasi Status Putusan Perkara

Tentunya informasi ini sangat bermanfaat bagi para pencari keadilan dan panmud perdata khusus untuk mengetahui persentase banyaknya gugatan kasasi untuk perkara perdata khusus dengan hasil perkara yang cepat dan lambat diputus.

Dalam eksperimen ini menggunakan dataset yang terdiri dari 8 atribut, datanya berupa numerik dan nominal. Metode yang diuji adalah algoritma klasifikasi decision tree, Naïve bayes, Logistic Regression, Decision Stump dan k-NN. Hasil eksperimen yang disajikan adalah *accuracy*, *precision*, *recall*, dan *AUC*.

Tabel 4.3 Hasil Eksperimen Algoritma Klasifikasi

| | | C4 | D | N | LR | k- |
|-------|-----|-----|-----|-----|-----|-----|
| | | .5 | S | B | | NN |
| Accur | acy | 96. | 48. | 91. | 93. | 95. |
| | AUC | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 |
| | | 84 | 36 | 34 | 72 | 00 |

Hasil eksperimen pada tabel diatas dapat dilihat bahwa perbandingan pengujian akurasi menunjukkan algoritma Decission Tree mempunyai tingkat akurasi terbaik dengan 96,02%, selanjutnya disusul oleh Decision Stump dengan 48.47%, Naive Bayes 91.80%, Logistic Regression 93.79% dan k-NN 95.67%. Sedangkan untuk pengujian AUC hasil terbaik yaitu Decission Tree sebesar 0,984. Jadi dapat diketahui Decission Tree keluar sebagai algoritma terbaik.

Tabel 4.4. Hasil Uji t-Test

| A | B | C | D | E | F |
|-------|-------|-------|-------|-------|-------|
| | B ↔ D | C ↔ D | D ↔ E | E ↔ F | F ↔ G |
| B ↔ D | | 0.993 | 0.900 | 0.998 | 0.990 |
| C ↔ D | | | | 0.998 | 0.990 |
| D ↔ E | | | | | 0.990 |
| E ↔ F | | | | | |
| F ↔ G | | | | | |

Keterangan:

B = Decision Tree (C4.5)

C = Naive Bayes (NB)

D = Logistic Regression (LR)

E = Decision Stump(DS)

F = k-Nearest Neighbour (k-NN)

Pada penelitian ini dilakukan pengujian dengan menggunakan uji *t-test*. Pada uji beda t-test nilai alpha yang digunakan adalah 0,05. Lima algoritma yang dilakukan uji beda t-test significance memiliki interpretasi yang berbeda, yaitu ada yang diterima atau ada perbedaan signifikan dan ada yang ditolak atau tidak ada perbedaan signifikan. Jika probabilitasnya <0,05 maka H0 ditolak dan jika >0,05 maka H0 diterima.

5. KESIMPULAN

Penelitian dengan menggunakan dataset register perkara perdata khusus dengan mengkomparasi 5 algoritma yaitu decision tree, Naïve bayes, Logistic Regression, Decision Stump dan k-NN. Dengan menggunakan validasi *10-fold cross validation* dan uji *t-test*. Hasil penelitian mendapatkan nilai akurasi tertinggi sebesar 96.02%

Tentunya informasi ini sangat bermanfaat bagi para pencari keadilan dan panmud perdata khusus untuk mengetahui persentase banyaknya gugatan kasasi untuk perkara perdata khusus dengan hasil perkara yang cepat dan lambat diputus.

Penelitian dilakukan dengan menerapkan konsep *business intelligence* dengan pengklasifikasian algoritma yang tepat dalam percepatan penyelesaian perkara dan sebagai alat bantu untuk mengolah dan menganalisa data register perkara perdata khusus sehingga dapat memberikan informasi yang akurat, sehingga panmud perdata khusus dapat memperkirakan jumlah perkara yang harus diselesaikan tepat waktu, dan penggunaan informasi tersebut membantu panmud perdata khusus dalam memanfaatkan metode *data mining* untuk percepatan penyelesaian perkara dalam membuat keputusan.

DAFTAR PUSTAKA

- Agung, M. (2014). Laporan tahunan mahkamah agung republik indonesia 2 0 1 4.
- Agung, M., & Indonesia, R. (2010). Cetak biru pembaruan peradilan 2010-2035.
- Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446–3453.

- <https://doi.org/10.1016/j.eswa.2011.09.033>
- Endres, A., & Rombach, D. (2003). A handbook of software and systems engineering, 346. <https://doi.org/10.1109/MS.2004.1270773>
- Fu, Y. (1997). Data mining. *IEEE Potentials*, 16(4), 18–20. <https://doi.org/10.1109/45.624335>
- Nurhadi. (2015). ROAD MAP 2015.pdf.
- Ranjan, J. (2009). Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*, 9, 60. <https://doi.org/10.2139/ssrn.2150581>
- Turban, E., Sharda, R., Delen, D., & King, D. (2011). Introduction to business intelligence. *Business Intelligence: A Managerial Approach*, 3–18. https://doi.org/10.1300/J155v11n01_01
- Undavia, J. N. (2014). Comparison of Classification Algorithms to Predict Comparison of Decision Tree Classification Algorithm to Predict Student $\hat{\epsilon}^M$ s Post Graduation Degree in Weka Environment, 1(2), 17–22.
- Vercellis, C. (2009). Business Intelligence: Data Mining and Optimization for Decision Making.
- Wahono, R. S., Herman, N. S., & Ahmad, S. (2014). A Comparison Framework of Classification Models for Software Defect Prediction, 20(10), 1945–1950. <https://doi.org/10.1166/asl.2014.5640>
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques Third Edition*. *Data Mining* (Vol. 277). USA: Morgan Kaufmann. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)
- Yu, L., Chen, G., Koronios, A., Zhu, S., & Guo, X. (n.d.). Application and Comparison of Classification Techniques in Controlling Credit Risk, 111–145.