

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pernikahan dalam pandangan Islam merupakan sesuatu yang sakral, bermakna ibadah kepada Allah SWT, mengikuti sunnah Rasulullah SAW dan dilaksanakan atas dasar keikhlasan, Tanggung Jawab, dan Mengikuti ketentuan-ketentuan hukum yang harus dindahkan. Sedangkan tujuan pernikahan sebagaimana difirmankan Allah SWT dalam Surat Ar-Rum ayat 21 yang artinya “Dan diantara Tanda-tanda Kekuasaan-Nya, ialah Dia menciptakan untukmu istri-istri dari jenismu sendiri, supaya kamu cenderung dan merasa tenteram kepadanya dan dijadikannya diantaramu rasa Kasih dan Sayang. Sesungguhnya pada yang demikian itu benar-benar terdapat tanda-tanda Bagi kaum yang berfikir”. Namun tidak dipungkiri perbedaan-perbedaan dalam pernikahan sering menimbulkan masalah ataupun pertengkaran antara suami istri. Munculnya permasalahan yang terjadi antara suami dan istri dapat terjadi karena kesalahpahaman, perselingkuhan, factor ekonomi, maupun faktor orang tua dan anak. Faktor-faktor diatas cenderung terjadi karena tidak ada penyeimbangan satu sama lain antara keduanya.[1]

Menurut laporan Statistik Indonesia, jumlah kasus perceraian di Indonesia mencapai 516.334 kasus pada 2022. Angka ini meningkat 15,31% dibandingkan 2021 yang mencapai 447.743 kasus. Jumlah kasus perceraian di Tanah Air pada tahun lalu bahkan mencapai angka tertinggi dalam enam tahun terakhir. Adapun mayoritas kasus perceraian di dalam negeri pada 2022 merupakan cerai gugat, alias perkara yang gugatan cerainya diajukan oleh pihak istri yang telah diputus oleh Pengadilan. Jumlahnya sebanyak 388.358 kasus atau 75,21% dari total kasus perceraian tanah air pada tahun lalu. Di sisi lain, sebanyak 127.986 kasus atau 24,78% perceraian terjadi karena cerai talak, yakni perkara yang permohonan cerainya diajukan oleh pihak suami yang telah diputus oleh pengadilan. Berdasarkan provinsinya, kasus perceraian tertinggi

pada 2022 berada di Jawa Barat, yakni sebanyak 113.643 kasus. Diikuti oleh Jawa Timur dan Jawa Tengah, masing-masing sebanyak 102.065 kasus dan 85.412 kasus. Terdapat lima provinsi yang tidak memiliki kasus perceraian sama sekali sepanjang 2022. Di antaranya Kepulauan Riau, Bali, Kalimantan Utara, Sulawesi Barat, dan Papua Barat. Adapun laporan tersebut mencatat, terdapat 448.126 perceraian di Indonesia yang terjadi berdasarkan faktor penyebabnya pada 2022. Penyebab utama perceraian pada 2022 adalah perselisihan dan pertengkaran. Jumlahnya sebanyak 284.169 kasus atau setara 63,41% dari total faktor penyebab kasus perceraian di tanah air. Kasus perceraian lainnya dilatarbelakangi alasan ekonomi, salah satu pihak meninggalkan, kekerasan dalam rumah tangga, hingga poligami.[2]

Setiap bulan di Kabupaten Kuningan, Pengadilan Agama mencatat setidaknya terdapat 250 hingga 300 perkara yang masuk ke pihaknya. Dari jumlah tersebut, pada bulan Juni 2022 mengalami peningkatan hingga 317 perkara. Sebanyak 317 perkara yang masuk Pengadilan Agama Kuningan selama bulan Juni 2022 diantaranya 212 cerai gugat atau perkara perceraian yang diajukan oleh istri, 60 cerai talak atau perceraian yang diajukan suami dan selebihnya merupakan hak asuh anak, penetapan ahli waris, dispensasi nikah, hingga hak asuh anak.[3]

Berdasarkan tingkat perceraian yang setiap tahun terus mengalami kenaikan, perlu adanya berupa data-data untuk mengetahui faktor-faktor yang mempengaruhi perceraian. Dengan demikian, data-data tersebut dapat dijadikan alat pengambilan keputusan yang nantinya dapat menjadi solusi baik bagi pemerintah, masyarakat maupun bagi pasangan suami istri dimasa yang akan datang. Salah satu teknik pengolahan data yang cocok dengan menggunakan teknik data mining. Salah satu teknik data mining yang dapat digunakan adalah *random forest*[1].

Data mining adalah sebuah kajian dari pengamatan untuk data dalam jumlah yang besar untuk mendapatkan suatu hubungan yang tidak diketahui sebelumnya dan ada dua metode baru untuk mempersingkat data agar mudah

dipahami serta berguna untuk pemilih data [4]. Selain menganalisa obesrvasi data dalam jumlah besar, data mining juga dapat menangani dimensi data yang tinggi dan juga data yang memiliki sifat berbeda[5].

Random forest adalah sebuah *algoritma ensemble* yang terdiri dari kumpulan pohon keputusan (decision trees). Setiap pohon keputusan dalam *random forest* melakukan prediksi secara independen, dan hasil prediksi dari setiap pohon digabungkan untuk menghasilkan prediksi akhir. Kelebihan dari *random forest* adalah kemampuannya dalam mengatasi masalah prediksi cuaca yang kompleks, termasuk non-linearitas, interaksi antar variabel, dan keberadaan outliers dalam data[6].

Proses pengajuan bantuan hukum dimulai ketika pemohon mengunjungi kantor Pusat Konsultasi dan Bantuan Hukum (PKBH). Di sana, pemohon menyerahkan dokumen atau berkas yang diperlukan untuk mendapatkan bantuan. Petugas di PKBH kemudian memeriksa kelengkapan berkas tersebut. Jika berkas dinyatakan lengkap, proses dapat dilanjutkan ke tahap berikutnya. Namun, jika berkas tidak lengkap, pemohon diminta untuk melengkapi dokumen yang diperlukan dan kembali ke PKBH untuk menyerahkannya. Setelah berkas dinyatakan lengkap, PKBH akan menentukan pengacara yang akan menangani kasus pemohon. Setelah pengacara ditentukan, pemohon melakukan konsultasi dengan pengacara tersebut untuk mempersiapkan kasusnya. Selanjutnya, pemohon didaftarkan ke pengadilan terkait, di mana pengadilan menetapkan jadwal sidang untuk kasus tersebut. Ketika jadwal sidang telah ditetapkan, pengacara akan menginformasikan tanggal dan waktu sidang kepada pemohon. Pada hari sidang, pengacara dan pemohon akan mendatangi pengadilan secara bersama-sama untuk menghadiri persidangan. Dengan mengikuti alur ini, pemohon mendapatkan bantuan hukum yang terstruktur mulai dari pengajuan hingga penjadwalan sidang di pengadilan, memastikan bahwa setiap langkah dilakukan sesuai prosedur yang berlaku dan pemohon memperoleh dukungan yang diperlukan dalam menghadapi permasalahan hukumnya.

Terdapat penelitian sebelumnya yang telah melakukan penelitian tentang putusan perceraian menggunakan data mining. Salah satu judul penelitian tersebut adalah "penerapan data mining dalam proses prediksi perceraian menggunakan *algoritma naive bayes* di kabupaten aceh tengah".[1] Penelitian ini bertujuan untuk menemukan faktor-faktor yang mempengaruhi perceraian dan menggunakan *algoritma naive bayes* sebagai alat pengambilan keputusan. Dengan asumsi bahwa kehadiran atau ketiadaan kejadian tertentu dalam suatu kelompok tidak berkorelasi, algoritma ini digunakan untuk memprediksi perceraian. Faktor-faktor seperti komunikasi pasif, ketidaksamaan, perselingkuhan, masalah nafkah batin, kesibukan pekerjaan berlebihan, kurangnya perhatian, pertengkaran, intimidasi, dan tindak kekerasan diidentifikasi sebagai kontributor perceraian. Penelitian ini menghasilkan sistem prediksi perceraian berbasis data mining dengan fokus pada pentingnya hukum islam dalam menetapkan standar perceraian. Sistem ini memberikan informasi berguna bagi pemerintah, masyarakat, dan pasangan untuk membuat keputusan terkait perceraian.

Penelitian lainnya yang telah melakukan penelitian tentang *random forest* menggunakan data mining. Salah satu judul penelitian tersebut adalah "prediksi stunting pada balita menggunakan *algoritma random forest*"[7]. Penelitian ini bertujuan untuk menggunakan metode *random forest* dalam memprediksi tingkat stunting pada balita dengan tingkat akurasi yang tinggi. Dengan menggunakan dataset yang terdiri dari 10,001 record data, 7 atribut, dan 1 kelas atribut, penelitian ini berhasil mencapai tingkat akurasi prediksi sebesar 85.86%. Hasil penelitian ini menunjukkan bahwa metode *random forest* merupakan pendekatan yang efektif dalam memprediksi tingkat stunting pada balita. Dengan demikian, penelitian ini memberikan kontribusi penting dalam upaya penanganan stunting pada balita, terutama dalam konteks Indonesia.

Dengan dilakukannya penelitian ini, diharapkan dapat bermanfaat bagi advokat dalam mengoptimalkan pemberian konsultasi hukum yang lebih efektif, dengan memahami berbagai faktor perceraian melalui analisis metode

random forest. Hasil penelitian diharapkan menjadi alat bantu dalam mediasi, memungkinkan advokat untuk merumuskan strategi negosiasi berdasarkan data dan kasus sebelumnya, serta mendorong pengembangan strategi pencegahan perceraian melalui kerja sama dengan kua, pengadilan, dan instansi terkait dalam penyelenggaraan program bimbingan pra-nikah. Selain itu, advokat diharapkan dapat memanfaatkan pengetahuan ini untuk edukasi masyarakat tentang pentingnya persiapan dan pemahaman sebelum menikah, serta mendorong kolaborasi lintas sektor untuk meningkatkan efektivitas program bimbingan pra-nikah dan strategi pencegahan perceraian secara komprehensif.

Berdasarkan uraian yang telah dipaparkan tersebut, maka peneliti tertarik untuk melakukan penelitian dalam bentuk skripsi dengan menentukan judul yaitu “Implementasi Metode Random Forest Untuk Menentukan Prediksi Putusan Perceraian Di Kabupaten Kuningan (Studi Kasus : PKBH Universitas Kuningan)”.

1.2 Identifikasi Masalah

Dari masalah-masalah yang muncul di latar belakang, maka dapat diidentifikasi masalah-masalah sebagai berikut :

1. Data statistik menunjukkan peningkatan signifikan dalam jumlah kasus perceraian di Indonesia, termasuk di kabupaten kuningan. Faktor-faktor penyebab perceraian sangatlah beragam, mulai dari perselisihan dan pertengkaran, masalah ekonomi, hingga kekerasan dalam rumah tangga.
2. Walaupun telah diidentifikasi beberapa faktor penyebab perceraian, masih diperlukan analisis mendalam mengenai seberapa signifikan pengaruh setiap faktor terhadap keputusan perceraian. Hal ini penting untuk memahami akar masalah yang sebenarnya dan mencari solusi yang tepat.

1.3 Rumusan Masalah

Berdasarkan uraian yang telah dijabarkan pada latar belakang mengenai permasalahan yang ada, maka rumusan masalah penelitian ini sebagai berikut:

1. Bagaimana tingkat efektivitas metode *random forest* dalam meramalkan putusan perceraian di Pengadilan Agama Kabupaten Kuningan.
2. Bagaimana tingkat akurasi prediksi perceraian yang diperoleh melalui penggunaan metode *random forest* di Pengadilan Agama Kabupaten Kuningan.
3. Bagaimana mengembangkan aplikasi prediksi perceraian berbasis metode *random forest* dapat memberikan kontribusi signifikan dalam membantu advokat di Kabupaten Kuningan untuk mengurangi tingginya angka perceraian.

1.4 Batasan Masalah

Agar pembahasan masalah menjadi lebih jelas dan terarah maka diperlukan adanya batasan masalah. Adapun ruang lingkup permasalahan dibatasi oleh:

1. Data yang digunakan tahun 2019-2023. Data yang diolah dan dianalisis melibatkan usia penggugat, usia tergugat, lama perkawinan, jumlah anak, faktor perceraian, dan hasil perceraian yang tercatat dalam Pengadilan Agama Kabupaten Kuningan.
2. Penelitian ini akan membatasi wilayahnya pada Kabupaten Kuningan sebagai contoh kasus, dengan tujuan untuk memahami dan menganalisis faktor-faktor yang mempengaruhi putusan perceraian.
3. Kualitas dan kelengkapan data historis tentang kasus perceraian di Kabupaten Kuningan sangat mempengaruhi akurasi model. Data yang tidak lengkap, tidak akurat, atau bias dapat menyebabkan prediksi yang tidak tepat.
4. Aplikasi ini hanya membatasi hak akses bagi admin dan personal.

1.5 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan tersebut, maka yang menjadi tujuan dalam penelitian ini adalah untuk merancang aplikasi prediksi putusan perceraian di Pengadilan Kuningan yang dapat digunakan untuk:

1. Mengevaluasi seberapa baik metode *random forest* dapat memprediksi keputusan perceraian di pengadilan agama kabupaten Kuningan dengan mempertimbangkan berbagai faktor penyebab.
2. Menentukan tingkat keakuratan ramalan perceraian yang dibuat oleh metode *random forest* di pengadilan agama kabupaten kuningan dengan melihat bagaimana keputusan perceraian yang sebenarnya dibuat dan ramalan yang dibuat oleh model.
3. Membangun aplikasi prediksi perceraian berbasis metode *random forest* untuk membantu advokat dan lembaga terkait di kabupaten kuningan dalam upaya mengurangi jumlah perceraian yang tinggi. Diharapkan aplikasi ini dapat membantu mengidentifikasi kasus-kasus yang berisiko tinggi terhadap perceraian agar intervensi dapat dilakukan lebih awal.

1.6 Manfaat Penelitian

1.6.1 Manfaat Teoritis

Secara teoritis penelitian ini memberikan manfaat pengetahuan mengenai implementasi metode *random forest* pada kasus prediksi putusan perceraian di kabupaten kuningan.

1.6.2 Manfaat Praktis

Hasil penelitian yang dilakukan oleh peneliti diharapkan pula dapat memberikan kegunaan atau manfaat praktis bagi beberapa pihak yaitu:

1. Bagi Peneliti

Penelitian ini diharapkan dapat menambah wawasan dan pengetahuan mengenai data mining dalam implementasinya pada aplikasi prediksi putusan perceraian.

2. Bagi Pengacara

Memberikan inovasi baru di bidang teknologi informasi berupa aplikasi prediksi putusan perceraian yang dapat memberikan informasi terkait tingkat akurasi prediksi perceraian.

3. Peneliti Lebih Lanjut

Penelitian ini diharapkan dijadikan peningkatan pengetahuan dan bahan referensi di bidang analisis data. Selain itu, pihak lain yang memerlukan informasi mengenai data mining, pembangunan aplikasi prediksi putusan perceraian, serta pengembangan sistem *agile*.

1.7 Pertanyaan Penelitian

Adapun pertanyaan penelitian yang digunakan dalam penelitian ini adalah:

1. Apakah metode *random forest* dapat meramalkan putusan perceraian di pengadilan agama kabupaten kuningan?
2. Apakah metode *random forest* dapat memprediksi kasus perceraian di kabupaten kuningan, dan apa variabel-variabel prediktor utama yang paling berpengaruh dalam melakukan prediksi tersebut?
3. Apakah signifikan pengaruh faktor-faktor seperti perselisihan, pertengkaran, masalah ekonomi, dan kekerasan dalam rumah tangga terhadap keputusan perceraian di kabupaten kuningan?
4. Apakah hasil prediksi perceraian akurat, relevan, dan aplikabilitas dalam konteks kasus-kasus spesifik di kabupaten kuningan?

1.8 Hipotesis Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, hipotesis penelitian yang diajukan yaitu:

1. Metode *random forest* sangat efektif dalam memprediksi keputusan perceraian di pengadilan agama kabupaten kuningan, metode ini dapat secara akurat mengidentifikasi komponen penting yang berkontribusi pada keputusan perceraian.

2. Metode *random forest* di pengadilan agama kabupaten kuningan membuat prediksi perceraian yang sangat akurat, ada banyak korelasi antara prediksi model dan keputusan perceraian yang sebenarnya.
3. Aplikasi prediksi perceraian yang dibuat dengan metode *random forest* membantu pengacara di kabupaten kuningan mengurangi jumlah perceraian. Ini karena aplikasi memberikan wawasan penting yang membantu dalam perencanaan dan pelaksanaan intervensi yang tepat waktu dan efektif.

1.9 Metodologi Penelitian

1.9.1 Metode Pengumpulan Data

a) Studi Pustaka

Studi pustaka ini dilakukan dengan menggunakan sumber-sumber seperti jurnal, buku dan internet. Studi literatur ini berguna untuk mengetahui landasan teori pengetahuan dan informasi pada penelitian, seperti jurnal referensi yang sesuai dengan judul penelitian yang diambil.

b) Observasi

Penelitian ini menggunakan observasi secara langsung di pengadilan agama kuningan. Observasi digunakan untuk mengetahui bagaimana keadaan kantor, dan ketersediaan data. Hasil dari observasi tersebut mendapatkan data primer.

c) Wawancara

Peneliti menggunakan teknik wawancara di pengadilan agama kabupaten kuningan untuk mengambil data primer terkait dengan penelitian.

1.9.2 Metode Pengembangan Sistem

1. *Agile*

Metode *agile* dipilih sebagai objek dalam penelitian ini karena metode *agile* merupakan salah satu metode pengembangan sistem yang cukup populer saat ini karena menawarkan fleksibilitas, sehingga

pengembang bisa kembali ke fase yang lebih awal apabila ada perubahan yang diperlukan. Metode *agile* adalah salah satu metodologi pengembangan perangkat lunak yang efektif dan tangkas.

Agile model awalnya dikembangkan karena pada metodologi tradisional terdapat banyak hal yang membuat proses pengembangan tidak dapat berhasil dengan baik sesuai tuntutan user. Saat ini metodologi ini sudah cukup banyak berkembang, di antaranya adalah *Xtreme Programming (XP)*, *Scrum Methodology*, *Crystal Family*, *Dynamic Systems Development Method (DSDM)*, *Adaptive Software Development (ASD)*. Dengan demikian *Agile* model tentunya memiliki kelebihan atau keunggulan dibandingkan dengan metode – metode yang lainnya.

Dengan fleksibilitas ini, pengguna dapat dengan cepat menyesuaikan diri dengan perubahan kebutuhan dan input, yang sangat penting untuk menganalisis data perceraian yang kompleks. Pendekatan iteratifnya menjamin algoritma *Random Forest* selalu diuji dan divalidasi, yang memastikan prediksinya akurat. Pengembang dan stakeholder bekerja sama untuk membuat aplikasi yang lebih sesuai dengan kebutuhan pengguna. Selain itu, *Agile* memudahkan pengiriman hasil awal dan memungkinkan penyesuaian cepat berdasarkan umpan balik. Metode ini meningkatkan efisiensi dan risiko sambil menjaga aplikasi tetap up-to-date dengan teknologi terbaru.



Gambar 1. 1 Agile Method Development

a. *Planning* (Perencanaan)

Bagian ini adalah tahapan pertama dari metode agile. Peneliti dan pihak staff IT pengelola data pengadilan membuat rancangan secara bersama sama. Untuk melakukan hal tersebut perlu dilakukan melalui komunikasi langsung berupa wawancara, dan mendapat data arsip dokumen yang dibutuhkan. Pada tahap ini peneliti setelah mendapatkan data terkait perceraian kemudian dilanjutkan dengan membuat desain dengan menggunakan tools system dan tools application.

b. *Implementation* (Implementasi)

Pada tahapan ini peneliti mengimplementasikan pengembangan sistem, tahapan dari perencanaan dengan membuat dokumentasi program dengan menggunakan rich picture, UML dan pembuatan user interface dengan menggunakan bahasa pemrograman PHP.

c. *Testing* (Tes Perangkat Lunak)

Untuk melakukan pengetesan program yang sudah dibuat apakah sudah sesuai atau belum. Pengujian tes dilakukan dengan berbagai cara diantaranya yaitu, pengujian black box dan pengujian white box.

d. *Documentation* (Dokumentasi)

Pada tahap ini dilakukan dokumentasi fungsi yang ada pada sistem informasi sebagai profil selama pengembangan dan proses yang dilakukan dalam dokumentasi antaralain fungsi-fungsi program, backup file-file yang digunakan pada saat pengembangan, rancangan database, subrutin, nilai input dan output.

e. *Deployment* (Penyebaran)

Tahap ini merupakan tahapan dalam pengembangan sistem dan menyediakan sistem bagi pengguna akhir. Yaitu Pengacara.

f. *Maintenance* (Pemeliharaan)

Pada titik ini, pemeliharaan sistem dilakukan secara berkala agar aman dari kegagalan sistem/kegagalan sistem, karena sistem belum tentu bebas dari kegagalan sistem.

1.9.3 Metode Penyelesaian Masalah

1. *KDD (Knowledge Discovery in Databases)*

Istilah Data Mining dan knowledge discovery in database (KDD) sering kali digunakan untuk mencari informasi yang tersembunyi pada sebuah database yang besar. KDD akan membantu untuk mengidentifikasi dan mengamati hal-hal yang ingin diketahui, dan mengolah data menjadi informasi dan berkaitan satu sama lain. Proses KDD salah satunya adalah Data Mining. Berikut proses KDD secara garis besar[8]:

a. Data Selection

Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada subset variable (sampel data) dimana penemuan (discovery) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional.

b. Pre-Processing dan Cleaning Data

Pre-Processing dan Cleaning data dilakukan membuang data yang tidak konsisten dan noise, duplikasi data, memperbaiki kesalahan data, dan bisa diperkaya dengan data eksternal yang relevan.

c. Transformation

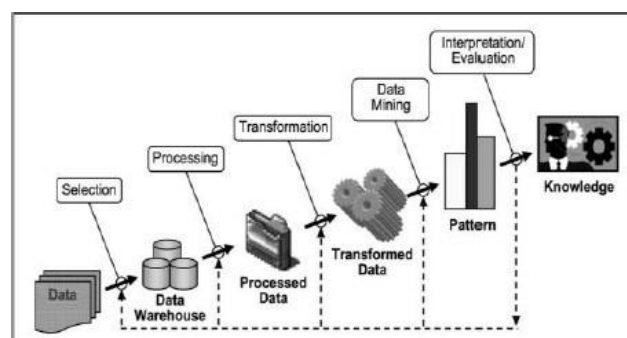
Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses mining dengan cara melakukan peringkasan (agregasi).

d. Data Mining

Proses Data Mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan.

e. Interpretation / Evaluasi

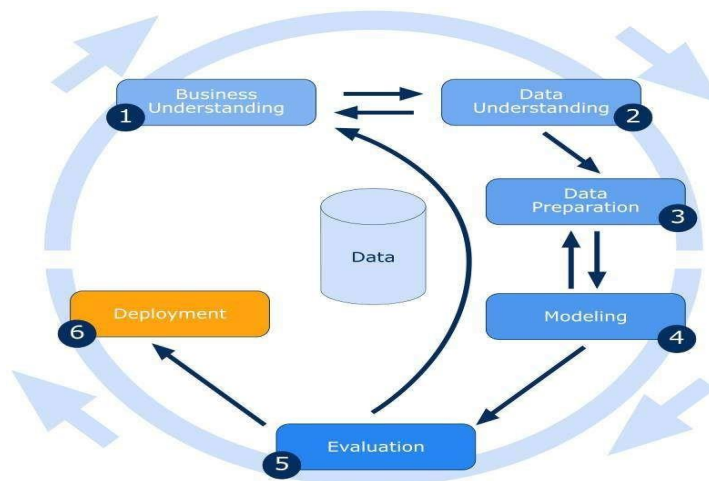
Proses untuk menerjemahkan pola-pola yang dihasilkan dari Data Mining. Mengevaluasi (menguji) apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta atau hipotesa sebelumnya. Pengetahuan yang diperoleh dari pola-pola yang terbentuk dipresentasikan dalam bentuk visualisasi [8]. Tahapan proses KDD dapat di lihat pada gambar :



Gambar 1. 2 Tahapan dalam KDD

2. CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM (Cross Industry Standard Process for Data Mining) suatu standarisasi pemrosesan data mining yang telah dikembangkan dimana data yang ada akan melewati setiap fase terstruktur dan terdefinisi dengan jelas dan efisien[9]. Selain menerapkan suatu model dalam proses penambangan data, pemilihan algoritma sangat mempengaruhi terhadap komparasi kinerja metode data mining. Metodologi ini terdiri dari enam tahapan yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Proses metodologi ini terdiri dari 6 tahapan yang dapat dijelaskan sebagai berikut.



Gambar 1. 3 Metodologi CRISP-DM

1. Business Understanding (Pemahaman Bisnis)

Beberapa hal yang dilakukan pada tahap ini seperti memahami kebutuhan serta tujuan dari sudut pandang bisnis selanjutnya mengartikan pengetahuan ke dalam bentuk pendefinisian masalah pada data mining dan kemudian menentukan rencana serta strategi untuk mencapai tujuan data mining.

2. Data Understanding (Pemahaman Data)

Tahapan ini diawali dengan mengumpulkan data, mendeskripsikan data, serta mengevaluasi kualitas data.

3. Data Preparation (Persiapan Data)

Dalam tahapan ini yaitu membangun dataset akhir dari berupa data mentah. Ada beberapa hal yang akan dilakukan mencakup melakukan pembersihan data (Data Cleaning), melakukan pemilihan data (Data Selection), record dan atribut-atribut, dan juga melakukan transformasi terhadap data (Data Transformation) untuk dijadikan masukan dalam tahap pemodelan.

4. Modelling (Pemodelan) Pada tahapan ini secara langsung melibatkan Machine Learning untuk penentuan teknik data mining, alat bantu data mining serta algoritma data mining. Penelitian ini menggunakan model klasifikasi Decision Tree dengan algoritma yang dipilih CART (Classification And Regression Tree). Cara algoritma CART untuk membentuk pohon keputusan dengan kriteria pemecah Gain Information [12]:

- Menentukan Akar/Root
- Menghitung Gain information masing-masing calon cabang
- Memilih atribut dengan Gain information yang mempunyai nilai terbesar sebagai splitting atribut yang akan dipilih sebagai cabang.
- Ulangi langkah 2,3 hingga terdapat leaf node

Gain information mengukur nilai impurity dari suatu partisi, A, dengan perhitungan:

$$\text{Gini}(D) = 1 - \sum_{i=1}^m p_i^2$$

Dimana:

Gini(D) = nilai *impurity* dari partisi D

m = jumlah indeks

p_i = peluang sebuah *tuple* D pada indeks ke i *Average Gini Impurity* dapat dihitung dengan.

$$Gini_A(D) = \frac{D_1}{D} Gini(D_1) + \frac{D_2}{D} Gini(D_2)$$

Dimana :

D = *tuple* D

D_1 = partisi pertama *tuple* D

D_2 = partisi kedua *tuple* D

$Gini_A(D)$ = impurity dari partisi D pada atribut A

$Gini(D_1)$ = impurity dari partisi pertama *tuple* D

$Gini(D_2)$ = impurity dari partisi kedua *tuple* D

Penurunan tingkat impurity yang diperoleh terhadap atribut A, bisa dihitung dengan :

$$\Delta Gini(A) = Gini(D) - Gini_A(D)$$

Dimana :

$\Delta Gini(A)$ = tingkat impurity

5. Evaluation (Pengujian)

Tahap ini dilakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh algoritma. Parameter yang digunakan untuk evaluasi komparasi algoritma adalah Confusion Matrix dengan aturan nilai akurasi, presisi dan recall. Nilai tersebut dapat diperoleh melalui perhitungan:

$$\text{Accuracy} = \frac{\text{Jumlah Klasifikasi Benar}}{\text{Total Sample testing yang diuji}} \times 100\%$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times 100\%$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \times 100\%$$

6. Deployment (Penyebaran)

Tahapan ini dilakukan dengan pembuatan laporan dan artikel jurnal menggunakan model yang dihasilkan[10].

3. *Random Forest*

Random forest merupakan metode klasifikasi yang dilakukan dengan mengembangkan metode *Decision Tree* berdasarkan pemilihan atribut acak pada setiap node untuk menentukan klasifikasi. Pada proses klasifikasinya didasarkan pada suara terbanyak dari pohon keputusan yang dikembalikan. *Random forest* merupakan gabungan dari pohon klasifikasi (*CART*) yang saling independen yang berasal dari distribusi yang sama melalui proses voting (jumlah terbanyak) untuk memperoleh prediksi klasifikasi. *Random forest* memiliki karakteristik yang dapat meminimumkan korelasi yang dapat menurunkan hasil kesalahan prediksi *random forest*. Ada tiga aspek penting dalam metode *random forest*, yaitu :

1. Melakukan bootstrap sampling untuk membangun pohon prediksi.
2. Masing-masing pohon keputusan memprediksi dengan prediktor acak. *Random Forest* membangun banyak pohon keputusan, dengan setiap pohon dilatih pada sampel data yang diambil secara acak dengan penggantian dari dataset asli (bootstrap sample). Selain itu, pada setiap titik pembagian (split) dalam pohon, hanya subset acak dari fitur yang dipertimbangkan untuk pembagian tersebut. Ini mengurangi korelasi antar pohon dalam forest dan meningkatkan kekuatan prediksi dari model secara keseluruhan. Untuk setiap pohon individu, rumus atau mekanisme pembagian (splitting) pada node-node tidak diatur oleh rumus khusus yang universal, melainkan oleh kriteria pemilihan split seperti Gini impurity atau entropy untuk klasifikasi, dan variance reduction untuk regresi, yang didefinisikan sebagai:

- Gini Impurity (untuk klasifikasi):

$$\text{Gini} = 1 - \sum_{i=1}^J P_i^2$$

- Entropy (alternatif untuk klasifikasi):

$$\text{Entropy} = - \sum_{i=1}^J P_i \log_2 (p_i)$$

- Variance Reduction (untuk regresi):

$$\text{Variance Reduction} = \text{Total Variance} - \sum_{j=1}^2 \frac{N_j}{N} \times \text{Variance}_j$$

3. Melakukan prediksi dengan mengkombinasikan hasil dari setiap pohon keputusan dengan cara majority vote untuk klasifikasi atau rata-rata untuk regresi.

Prediksi kelas = modus(prediksi dari semua pohon)

Rumus random forest :

$$\text{Prediksi} = \frac{1}{N} \sum_{i=1}^N \text{Prediksi Pohon}_i$$

di mana N adalah jumlah pohon dalam forest.

Keterangan :

Contoh Soal:

Dataset:

Misalkan kita memiliki dataset sebagai berikut:

$$D = \{(30, 32, 1, 0, 60, 0), (35, 37, 2, 1, 72, 1), (28, 30, 0, 0, 48, 0)\}$$

Dengan Keterangan:

- Usia Penggugat
- Usia Tergugat
- Jumlah Anak
- Faktor Perceraian (0: Tidak, 1: Ya)
- Lama Perkawinan/Pisah Rumah (dalam bulan)
- Keputusan Cerai atau Tidak (0: Tidak cerai, 1: Cerai)

1. Penyusunan Dataset:

Dataset ini mencakup beberapa pasangan yang ingin diprediksi apakah akan bercerai atau tidak. Setiap baris merepresentasikan satu pasangan dengan fitur-fitur tertentu.

2. Proses Bootstrap:

Manfaat, Metodologi Penelitian, Teknik Pengumpulan Data, Lokasi Penelitian, dan Sistematika Penulisan.

BAB II : LANDASAN TEORITIS

Bab ini berisi mengenai teori terkait penelitian, menguraikan teori yang digunakan, peneliti sebelumnya, dan kerangka teoritis. Landasan teori meliputi tentang pengertian Data Mining, pengenalan Random Forest, Prediksi, pemodelan sistem, dan Perangkat Lunak yang digunakan.

BAB III : ANALISA DAN PERANCANGAN

Bab ini berisi perancangan yang dibuat menggunakan Rich picture, UML, dan komponen lainnya. Bab ini juga berisi desain sistem yang akan dibangun.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Bab pengujian dan implementasi ini, membahas hasil-hasil dari tahapan penelitian, analisis, desain, implementasi desain, hasil pengujian dan implementasi. Hasil pengujian (*testing*) program menggunakan pengujian kotak hitam (*black-box testing*) dan kotak putih (*white-box testing*).

BAB V : KESIMPULAN DAN SARAN

Bab penutup ini membahas mengenai kesimpulan dari laporan penelitian yang dibuat serta saran untuk pengembangan program dari Pengadilan Kuningan.