**SISTEM REKOMENDASI KELAYAKAN KREDIT MENGGUNAKAN METODE RANDOM FOREST PADA BRI KANTOR CABANG PELAIHARI**

**PROPOSAL SKRIPSI**



**DISUSUN OLEH:**

**MUHAMMAD IRHAMNA PUTRA H76215022**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS ISLAM NEGERI SUNAN AMPEL**

**SURABAYA**

**2019**

# LEMBAR PERSETUJUAN PEMBIMBING

|  |  |  |
| --- | --- | --- |
| JUDUL | : | SISTEM REKOMENDASI KELAYAKAN KREDIT MENGGUNAKAN METODE RANDOM FOREST PADA BRI KANTOR CABANG PELAIHARI |
| NAMA | : | MUHAMMAD IRHAMNA PUTRA |
| NIM | : | H76215022 |

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 01 Oktober 2019.

Menyetujui,

|  |  |
| --- | --- |
| Dosen Pembimbing I | Dosen Pembimbing II |
|  |  |
|  |  |
|  |  |
|  |  |
| (Ahmad Yusuf, M. Kom) | (Nita Yalina, S.Kom., M.MT) |
| NIP. 199001202014031003 | NIP. 198702082014032003 |

|  |
| --- |
| Ketua Program Studi |
| Sistem Informasi |
|  |
|  |
|  |
| (Muhammad Andik Izzudin, M.T) |
| NIP. 19849397291431001 |

# LEMBAR PENGESAHAN

|  |  |  |
| --- | --- | --- |
| JUDUL | : | SISTEM REKOMENDASI KELAYAKAN KREDIT MENGGUNAKAN METODE RANDOM FOREST PADA BRI KANTOR CABANG PELAIHARI |
| NAMA | : | MUHAMMAD IRHAMNA PUTRA |
| NIM | : | H76215022 |

Proposal skripsi tersebut telah dipresentasikan pada Sidang Proposal Skripsi di depan Dosen Penguji pada tanggal 02 Mei 2019.

Menyetujui,

|  |  |
| --- | --- |
| Dosen Pembimbing I | Dosen Pembimbing II |
|  |  |
|  |  |
|  |  |
|  |  |
| (Ahmad Yusuf, M. Kom) | (Nita Yalina, S.Kom., M.MT) |
| NIP. 199001202014031003 | NIP. 198702082014032003 |
|  |  |
| Dosen Penguji I | Dosen Penguji II |
|  |  |
|  |  |
|  |  |
|  |  |
| (Mujib Ridwan, S.Kom., M.T) | (Nurissaidah Ulinnuha, M. Kom) |
| NIP. 198604272014031004 | NIP. 199011022014032004 |

Mengetahui,

|  |  |
| --- | --- |
| Ketua program Studi | Ketua jurusan |
|  |  |
|  |  |
|  |  |
|  |  |
| (Muhammad Andik Izzuddin, MT) | (Mujib Ridwan, S.Kom., M.T) |
| NIP. 198403072014031001 | NIP. 198604272014031004 |

# PERNYATAAN KEASLIAN

Saya yang bertandatangan di bawahini:

|  |  |  |
| --- | --- | --- |
| Nama | : | Muhammad Irhamna Putra |
| NIM | : | H76215022 |
| Program Studi | : | Sistem Informasi |
| Angkatan | : | 2015 |
|  |  |  |

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: “SISTEM REKOMENDASI KELAYAKAN KREDIT MENGGUNAKAN METODE RANDOM FOREST PADA BRI KANTOR CABANG PELAIHARI”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 02 Oktober 2019

Yang menyatakan,

(Muhammad Irhamna Putra)

NIM. H76215022

# HALAMAN MOTTO

“Dan pada akhirnya, jutaan *quotes* tidaklah bermakna jika yang ada hanyalah halusinasi tanpa eksekusi”

# HALAMAN PERSEMBAHAN

Segala puji bagi Allah SWT, atas segala nikmat, karunia serta ridho-Nya yang senantiasa menuntun langkahka dan memberiku kekuatan dalam menyelesaikan skripsi ini. Tak lupa, Sholawat serta salam saya haturkan pada junjungan Nabi besar umat Islam, Nabi Muhammad SAW.

Saya persembahkan karya ini untuk kedua orang tua tercinta saya, Ayahanda Syaifuddin, dan Ibunda Maisyarah yang telah memberikan do’a, motivasi dan nasihat yang tiada hentinya yang selalu memberi dukungan, motivasi, kasih sayang, dan telah sabar atas semua kelakuanku.

Semua pahlawan tanda jasa, guru dan dosen dari TK hingga Perguruan Tinggi. Bapak dosen pembimbing skripsi, Bapak Ahmad Yusuf, M. Kom dan Ibu Nita Yalina, S.Kom., M.MT, yang telah tulus dan sabar mendidik, membimbing, dan memberikan ilmunya kepada kami.

Teman-teman EXIST’15 yang tak dapat saya sebutkan satu per satu, Sobat Ambyar / Noto Urip Ben Tuo Ne Penak / DISC’15. Terimakasih atas kerjasama, kenangan, kebersamaan, dan pengalaman bersama kalian tak akan terlupakan dan tergantikan.

# KATA PENGANTAR

Segala puji syukur penulis panjatkan kehadirat Allah SWT, atas segala nikmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Sistem Rekomendasi Kelayakan Kredit Menggunakan Metode Random Forest Pada Bri Kantor Cabang Pelaihari”.** Skripsi ini penulis susun sebagai salah satu syarat untuk menyelesaikan studi Strata I di Program Studi Sistem Informasi, Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Ampel Surabaya.

Skipsi ini penulis susun dari berbagai acuan melalui studi pustaka pada beberapa buku-buku diktat, dan jurnal-jurnal pendukung yang dicuplik berdasarkan aturan-aturan cuplikan karya ilmiah secara benar dan dapat dipertanggungjawabkan. Adapun dengan adanya skripsi ini, penulis berharap hasilnya dapat digunakan sebagai referensi bagi adik-adik maupun orang lain sebagai tambahan ilmu pengetahuan.

Skripsi ini terselesaikan melalui kerja keras, niat dan kesungguhan penulis serta bantuan dan dukungan dari berbagai pihak berupa fisik maupun materil. Pada kesempatan ini tidak lupa penulis menyampaikan rasa terimakasih sebesar-besarnya kepada:

1. Allah SWT atas segala nikmat dan kuasa-Nya sehingga penulis dapat merantau dan menimba ilmu di Universitas Islam Negeri Sunan Ampel Surabaya ini untuk bekal dunia dan akhirat.
2. Ayahanda Syaifuddin dan Ibunda Maisyarah, selaku orang tua penulis yang telah memberikan seluruh kebaikan dalam kehidupannya, serta do’a, motivas dan nasihat tanpa henti.
3. Bapak Andik Izzuddin, M.T selaku Ketua Prodi Sistem Informasi Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya atas bantuan, motivasi, dan bimbingan yang diberikan.
4. Bapak Ahmad Yusuf, M. Kom dan Ibu Nita Yalina, S.Kom., M.MT, selaku Pembimbing yang telah menyediakan waktu, tenaga dan pikiran untuk mengarahkan penulis dalam penyusunan skripsi ini.
5. Adik tercinta penulis, Irsyad Musyaffa dan Irma Hanania yang telah memberikan semangat dan dorongan yang tiada henti.
6. Sahabat-sahabat EXIST’15 yang telah banyak membantu, menyemangati, memberikan dukungan, semangat, dan setia menemani dalam penulisan skripsi ini.
7. KKN 68 Desa Tawangrejo, yang telah memberikan pengalaman serta pelajaran berharga dalam kehidupan ini.
8. Yang diharapkan untuk selalu menemani penulis dikehidupan mendatang, skripsi ini takkan tertulis dan teredit dengan sempurna tanpa bantuannya, Rif’atul Farizati Nabilah.
9. Semua pihak yang telah membantu dalam penyelesaian skripsi ini.

Penulis mengakui atas kekurangan-kekurangan yang ada dalam skripsi, dan bahkan mungkin masih jauh dari kesempurnaan. Untuk itu penulis menerima dengan senang hati segala kritik dan saran yang bersifat membangun dari pembaca, sehingga penelitian selanjutnya diharapkan bias lebih baik dan semoga skripsi ini dapat bermanfaat bagi semua pihak, serta perkembangan ilmu pengetahuan, dan dapat berkontribusi terhadap kemajuan UINSA, bangsa, dan Negara. Akhir kata, penulis sampaikan terimakasih kepada semua pihak yang telah membantu menyelesaikan skipsi ini dan semoga skipsi penelitian ini bermanfaat.

Surabaya, 02 Oktober 2019

**Muhammad Irhamna Putra**

**NIM. H76215022**

# DAFTAR ISI

[LEMBAR PERSETUJUAN PEMBIMBING i](#_Toc20876299)

[LEMBAR PENGESAHAN ii](#_Toc20876300)

[KATA PENGANTAR iii](#_Toc20876301)

[DAFTAR ISI v](#_Toc20876302)

[BAB I PENDAHULUAN vi](#_Toc20876303)

[1.1. Latar Belakang vi](#_Toc20876304)

[1.2. Perumusan Masalah viii](#_Toc20876305)

[1.3. Batasan Masalah ix](#_Toc20876306)

[1.4. Tujuan Penelitian ix](#_Toc20876307)

[1.5. Manfaat Penelitian ix](#_Toc20876308)

[1.6. Sistematika Penulisan Skripsi x](#_Toc20876309)

[BAB II TINJAUAN PUSTAKA xi](#_Toc20876310)

[2.1. Tinjauan Penelitian Terdahulu xi](#_Toc20876311)

[2.2. Dasar Teori xiii](#_Toc20876312)

[2.2.1. Sistem Pendukung Keputusan (SPK) xiii](#_Toc20876313)

[2.2.2. Klasifikasi xiv](#_Toc20876314)

[2.2.3. Algoritma Random Forest xiv](#_Toc20876315)

[2.2.4. *Confusion Matrix* dan *Accuracy* xviii](#_Toc20876316)

[2.2.5. Analisa Kredit xix](#_Toc20876317)

[2.2.6. Database xx](#_Toc20876318)

[2.3. Integrasi Keilmuan xx](#_Toc20876319)

[BAB III METODOLOGI PENELITIAN xxiv](#_Toc20876320)

[3.1. Desain Penelitian xxiv](#_Toc20876321)

[3.1.1. Perumusan Masalah xxv](#_Toc20876322)

[3.1.2. Studi Pustaka xxv](#_Toc20876323)

[3.1.3. Pengumpulan Data xxv](#_Toc20876324)

[3.1.4. Pengolahan Data xxvii](#_Toc20876325)

[3.1.5. Pengembangan Sistem xxviii](#_Toc20876326)

[3.1.6. Pengujian Sistem xxviii](#_Toc20876327)

[BAB IV HASIL DAN PEMBAHASAN xxx](#_Toc20876328)

[4.1.1. Pengenalan algoritma *Random Forest* xxx](#_Toc20876329)

[4.1.2. *Preprocessing data* xxxiv](#_Toc20876330)

[4.1.3. Implementasi *Random Forest* xxxvi](#_Toc20876331)

[4.2. Pengembangan sistem xxxviii](#_Toc20876332)

[4.2.1. Analisis kebutuhan sistem xxxviii](#_Toc20876333)

[4.2.2. Desain sistem xxxix](#_Toc20876334)

[4.2.3. Implementasi xlii](#_Toc20876335)

[5.1.1. *Testing* xlvi](#_Toc20876336)

[5.1.2. Perawatan xlvii](#_Toc20876337)

[4.3. Pengujian Sistem xlvii](#_Toc20876338)

[BAB V HASIL DAN PEMBAHASAN l](#_Toc20876339)

[5.1. Kesimpulan l](#_Toc20876340)

[5.2. Saran l](#_Toc20876341)

[DAFTAR PUSTAKA li](#_Toc20876342)

# BAB I PENDAHULUAN

## Latar Belakang

Menjadi salah satu lembaga untuk memberikan pelayanan terhadap jasa intermediasi keuangan, Bank yang kemudian diatur menurut UU No 10 Tahun 1998 tanggal 10 November 1998 tentang perbankan, jasa atau usaha perbankan memiliki tiga kegiatan utama, yaitu menghimpun dana, menyalurkan dana serta memberikan jasa bank lainnya. Dalam kegiatan menyalurkan dana lebih tepatnya dalam pemberian kredit, ada sebuah istilah resiko kredit yang wajib dikelola oleh bank manapun. Resiko kredit ini adalah kegagalan atau ketidakmampuan dari debitur untuk memenuhi kewajibannya dalam hal melakukan pembayaran sesuai atas ketentuan yang telah disepakati sebelumnya. Dalam hal ini, tentu saja pihak Bank wajib memerlukan ketelitian dengan tingkat tinggi dalam memilih calon debitur dalam rangka mengurangi resiko kredit. Dalam resiko kredit, terdapat sebuah istilah kredit macet yang mana merupakan masalah utama dalam resiko kredit yang menjadi satu masalah yang kompleks. Kurang tepatnya melakukan penilaian awal pada calon debitur merupakan penyebab dari masalah ini.

Bank Rakyat Indonesia atau kerap dikenal sebagai Bank BRI tercatat sebagai bank yang paling besar meraup untung pada tahun 2017, berfokus pada penyaluran kredit ke usaha mikro, kecil dan menengah (UMKM) berhasil mencetak laba sebesar Rp. 13,4 triliun atau naik sekitar 10,4% dibanding periode tahun sebelumnya. Adapun pertumbuhan laba ini utamanya terdorong oleh tingginya penyaluran kredit. Total penyaluran kredit BRI sepanjang tahun 2017 tercatat sebesar Rp 687,9 triliun atau naik 11,8%. Terkait dengan rasio kredit bermasalah atau *Non Performing Loan* (NPL), pada tahun 2017 dengan pertumbuhan kredit yang signifikan tersebut, BRI berhasil menjaga rasio NPL gross nya sebesar 2,16% atau turun dibandingkan dengan NPL gross pada tahun 2016 sebesar 2,22%. Namun peningkatan tersebut merupakan statistik penurunan NPL secara garis besar, faktanya dalam lapangan ada beberapa cabang yang masih memiliki peningkatan NPL misalnya Bank BRI Kantor Cabang Pelaihari. Dalam wawancara secara pribadi, disebutkan bahwa NPL pada kantor cabang tersebut mengalami peningkatan yang nilainya tidak boleh disebutkan. Hal ini dikarenakan pertumbuhan ekonomi di masyarakat yang mempengaruhi terhadap nilai NPL ini. Maka dari itu, untuk menekan peningkatan nilai NPL ini diharuskan untuk mengelola kelayakan kredit secara lebih mendetail lagi untuk calon nasabah atau debitur baru.

Kelayakan kredit merupakan kriteria penentu terhadap layak atau tidaknya calon debitur dalam hal diberikan kredit agar tidak menimbulkan kemacetan kredit yang berimbas pada rasio kredit bermasalah atau *Non Performing Loan* (NPL). Kelayakan kredit tersebut dapat diukur dengan skor kredit yang merupakan angka yang diberikan kepada suatu individu yang menjelaskan seberapa besar kemungkinan individu tersebut dalam memenuhi kewajiban finansialnya. Hal ini dapat di lihat sebagai probabilitas standar untuk mengukur calon debitur. Pemberi pinjaman, terutama institusi bank atau Lembaga lainnya memperhitungkan kemungkinan standar ini untuk menetapkan calon debitur apakah pinjaman ini nantinya akan masuk ke dalam kategori pinjaman baik (*Good Loan* atau pinjaman yang memiliki kemungkinan untuk membayar kembali pinjaman) atau pinjaman buruk (*Bad Loan* atau pinjaman yang tidak memiliki kemungkinan untuk membayar kembali pinjaman) berdasarkan probabilitas standar tersebut, sehingga individu tersebut dapat diklasifikasikan ke kategori yang telah ada. Atas dasar hal ini, dibutuhkanlah sebuah sistem pendukung keputusan untuk meninjau calon debitur yang memiliki potensi dan layak diberikan kredit.

Sistem Pendukung Keputusan adalah sistem berbasis komputer yang memiliki tujuan untuk membantu dalam pengambilan keputusan dengan memanfaatkan data dan model tertentu untuk memecahkan berbagai persoalan yang tidak terstruktur. Secara umum sistem pendukung keputusan memiliki peran untuk membantu pengambilan keputusan secara efektif. Sistem pendukung keputusan dengan menggunakan dan memanipulasi data sedemikian rupa, lalu menampilkannya dengan memberikan *user interface* yang *easy learning*, sehingga dapat menggabungkan pemikiran pengambil keputusan dari proses tersebut.

Metode Random Forest merupakan metode yang cukup efektif yang mampu memberikan rekomendasi terbaik dibandingkan dengan metode – metode *machine learning* lainnya. *Random Forest* merupakan salah satu metode atau teknik untuk kebutuhan klasifikasi *data mining* yang mana termasuk dalam metode *ensemble.* Dikatakan metode *ensemble* dikarenakan metode *random forest* ini merupakan metode yang tercipta atas beberapa gabungan dari pohon keputusan pohon keputusan, sehingga menghasilkan sebuah keputusan dari gabungan pohon tersebut. Metode *random forest* ini berasal dari pengembangan metode *Classification and Regression Tree* (CART), yakni dengan menerapkan metode *bootstrap aggregating* (*Bagging*) dan *random feature selection* (Breiman, 2001). Metode *random forest* ini telah diaplikasikan pada berbagai permasalahan dalam situasi penelitian yang berkecimpung pada kesehatan, bisnis, Pendidikan dan lainnya.

Sehubungan dengan pesatnya perkembangan teknologi di era ini, diperlukan sebuah sistem pendukung keputusan dengan cara menganalisa data rekapan calon debitur yang telah atau tidak diberikan pinjaman yang ada pada tahun sebelumnya. Pada tahap selanjutnya sistem akan menghitung kelayakan calon debitur baru berdasarkan data rekapan tersebut. Dengan harapan sistem ini kedepannya akan membantu pihak bank dalam mengurangi resiko kredit.

## Perumusan Masalah

Berdasarkan latar belakang yang terpaparkan diatas, dapat ditarik rumusan masalah sebagai berikut.

1. Bagaimana mengembangkan aplikasi penentuan kelayakan kredit?
2. Bagaimana evaluasi metode *random forest*?

## Batasan Masalah

Adapun Batasan – Batasan masalah agar penelitian ini tidak melebar dari cakupan masalah dalam penelitian ini adalah:

1. Data penelitian yang akan digunakan didapatkan dari Kantor BRI Cabang Pelaihari, Kalimantan Selatan.
2. Data yang digunakan merupakan data pinjaman berdasarkan pinjaman yang berlangsung pada tahun 2015 – 2019.
3. Sistem pendukung keputusan ini akan dibangun berbasis *web*.

## Tujuan Penelitian

1. Dapat mengembangkan aplikasi penentuan kelayakan kredit dengan menggunakan metode Random Forest.
2. Dapat membuktikan bahwa metode Random Forest mampu mengurangi resiko kredit pada Bank.

## Manfaat Penelitian

Hasil dari penelitian ini adalah diharapkannya mampu memberikan manfaat untuk semua pihak yang terlibat dalam penelitian ini, diantaranya:

1. Bagi Peneliti
   1. Sebagai media dalam menerapkan ilmu pengetahuan yang telah diperoleh selama melaksanakan kuliah.
   2. Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer Program Studi Sistem Informasi di Universitas Islam Negeri Sunan Ampel Surabaya.
2. Bagi Pembaca
   1. Sebagai referensi ilmu pengetahuan dalam pengembangan aplikasi perhitungan dan metode Random Forest.
   2. Membuka wawasan dalalm pengembangan sebuah sistem dengan menggunakna metode Random Forest.
3. Bagi Bank
   1. Dengan adanya aplikasi penentuan kelayakan kredit, mampu mengurangi resiko kredit dalam menentukan calon debitur baru.
   2. Memberikan penilaian terhadap calon debitur baru yang berhak menerima kredit dari pihak bank sesuai dengan keadaan sebenarnya.
   3. Digunakan sebagai dasar pengambilan keputusan yang terstruktur untuk mendukung proses utang piutang pada Bank BRI Cabang Pelaihari.
   4. Dapat digunakan untuk penentuan kelayakan pemberian kredit pada debitur.
   5. Menjadi jembatan antara Calon Debitur Baru dan Pihak Bank dalam proses pengajuan kredit.

## Sistematika Penulisan Skripsi

Adapun sistematika penulisan skripsi ini terdiri atas beberapa bagian, yakni sebagai berikut:

1. Bagian awal terdiri dari halaman sampul.
2. Bagian isi terdiri dari :
   1. BAB I PENDAHULUAN

Pendahuluan menjelaskan latar belakang permasalahan yang diangkat dalam penelitian, perumusan masalah, tujuan penelitian, dan manfaat yang diperoleh dari penelitian yang dilakukan.

* 1. BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang penjelasan penelitian sebelumnya yang relevan dengan penelitian ini, dan penjelasan singkat tentang teori – teori yang terkait dengan penelitian ini seperti konsep sistem informasi, metode yang terkait dengan penelitian ini seperti konsep sistem informasi, metode pengembangan sistem informasi, Bahasa pemrograman, metode random forest dan sebagainya.

* 1. BAB III METODOLOGI PENELITIAN

Bab ini berisikan alur sistematika penelitian yang terdiri dari tahap identifikasi kebutuhan, tahap perencanaan dan tahap pengembangan prototipe dan tahap implementasi.

* 1. BAB IV HASIL DAN PEMBAHSAN

Bab ini berisikan hasil dan pembahasan bagaimana data diolah berdasarkan algoritma *Random Forest* dan bagaimana suatu sistem dikembangkan.

* 1. BAB V PENUTUP

Bab ini berisikan kesimpulan dari analisa masalah serta saran – saran dari penulis, sehingga apa yang menjadi tujuan dari penelitian ini dapat terwujud.

# BAB II TINJAUAN PUSTAKA

## Tinjauan Penelitian Terdahulu

Dalam memberikan pemahaman mengenai keterkaitan penelitian terdahulu dengan penelitian yang akan dilakukan, dapat dilihat lebih lanjut dalam Tabel 2.1:

**Tabel 2.1**: Daftar Penelitian Terdahulu

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Topik** | **Metode** | **Hasil** |
| 1. | KLASIFIKASI PEMBIAYAAN WARUNG MIKRO MENGGUNAKAN METODE *RANDOM FOREST* DENGAN TEKNIK *SAMPLING* KELAS *IMBALANCED* (Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi) (Widiastuti, 2018) | *Random Forest* | Metode *Random Forest* digunakan pada *Imbalanced Data* sehingga ketika di *training* akurasi mencapai 82.54%, namun nilai *sensitivity* bernilai 0, sehingga akurasi dianggap bias. Hasil klasifikasi didapatkan dengan nilai estimasi error OOB sebesar 25.44% |
| 2. | SISTEM PENDUKUNG KEPUTUSAN KELAYAKAN PEMBERIAN KREDIT PADA BANK TABUNGAN NEGARA (BTN) MENGGUNAKAN ALGORITMA C4.5 (Octabriyantiningtyas, 2016) | *Decision Tree* – C4.5 | Sistem dievaluasi dan dikategorikan baik, sehinggap dapat disimpulkan bahwa sistem bersifat *user friendly*.  Berdasarkan hasil penelitian, tingkat akurasinya rule – rule dalam sisitem ini melebihi 50%. Sehingga dapat digunakan sebagai dasar pengembangan sistem pendukung keputusan untuk Analisa pemberian kredit. |
| 3. | SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN KREDIT KOPERASI SERBA USAHA BERKAH TIRAM JAYA 4MENGGUNAKAN METODE *ANALYTIC HIERARCHY PROCESS* (AHP) (Nurdiyanto and Minarto, 2017) | * *Analytic Hierarchy Process* (AHP) | Sistem dapat membantu pihak koperasi untuk menentukan penerima kredit dengan mengambil nilai hasil akhir nasabah dengan prioritas tertinggi. |
| 4. | SISTEM PENDUKUNG KEPUTUSAN KREDIT USAHA RAKYAT PT. BANK RAKYAT INDONESIA UNIT KALIANGKRIK MAGELANG (Nugroho, Kusrini and Arief, 2015) | *K-Nearest Neighbors* (KNN) | Sistem dapat digunakan untuk menentukan baik, buruk, normal calon penerima kredit KUR. Penerapan algoritma KNN diperoleh tingkat *error rate* sebesar 6,98% dan 93,023% akurat. |
| 5. | Klasifikasi Nasabah Thera Bank Membeli *Personal Loan* Menggunakan Metode Klasifikasi Dalam *Machine Learning* (Putri, 2018) | * *Random Forest* * *Naives-Bayes Classifier* * *K-Nearest Neighbors* * *Decision Tree* * *Random Forests* | Dengan membandingkan beberapa metode klasifikasi dalam *machine learning*, didapatkan bahwa metode terbaik adalah *random forest* dengan *feature selection* dan *hyperparameter tuning*. |
| 6. | Kategorisasi Teks pada Hadits Sahih Al-Bukhari menggunakan *Random Forest* (Afianto *et al.*, 2017) | *Random Forests* | Kategorisasi dokumen hadits sahih Al-Bukhari dengan menggunakan metode *Random Forests* dengan mekanisme *Preprocessing* berupa *Stemming, Case Folding* dan *Filtration* menghasilkan nilai *F1-Score* sebesar 90%. |
| 7. | PREDIKSI LAMA STUDI MAHASISWA DENGAN METODE RANDOM FOREST (Budi Adnyana, 2016) | *Random Forests* | Hasil pengujian menunjukkan nilai *accuracy* sebesar 83.54%. |
| 8. | An up-to-date comparison of state-of-the-art classification algorithms (Zhang *et al.*, 2017) | 11 Algoritma | Dengan membandingkan 11 algoritma populer dalam machine learning, didapatkan 5 algoritma yang telah di urutkan berdasarkan peringkat tinggi akurasinya, yakni:   1. GBDT 2. *Random Forests* 3. SVM 4. C4.5 |

Dari beberapa matrik penelitian diatas dapat dikatakan bahwa saat ini banyak penelitian terhadap analisis hingga klasifikasi terhadap kelayakan kredit, sistem pendukung keputusan serta penggunaan *machine learning* untuk keperluan klasifikasi hingga *data mining* dari berbagai macam akademisi.

Adapun penggunaan algoritma *random forest* termasuk sebagai salah satu algoritma yang populer yang penggunaannya digunakan dalam hal *classification* dan *clustering*. Berdasarkan penelitian sebelumnya, *random forest* memiliki tingkat akurasi yang tinggi. Sehingga, peneliti tertarik untuk menggunakan *random forests* dalam klasifikasi terhadap analisa kelayakan kredit pada bank.

## Dasar Teori

Dasar teori berisi teori yang relevan yang mencakup berbagai perihal tentang variabel yang akan diteliti.

## Sistem Pendukung Keputusan (SPK)

Sistem pendukung keputusan adalah sebuah sistem yang dibangun untuk mendukung kebutuhan dalam hal pengambilan keputusan dalam suatu organisasi perusahaan yang dirancang guna pengembangan efektifitas dan produktifitas para eksekutif dalam hal ini bagian manajerial dengan bantuan teknologi komputer (Manurung, 2010)

Sistem keputusan merupakan sebuah sistem yang terbentuk dari dua buah model sistem, yakni keputusan bersifat tertutup dan keputusan bersifat terbuka. Sistem keputusan bersifat tertutup memiliki arti bahwa keputusan dipisah dari masukan yang tidak diketahui lingkungannya, sedangkan sistem keputusan bersifat terbuka memandang bahwa keputusan merupakan bagian dari suatu lingkungan yang kompleks dan juga merupakan dari bagian pihak – pihak yang tidak diketahui dalam hal ini, dilihat dari berbagai aspek yang ada dilingkungannya. Sebuah keputusan dapat dihasilkan karena dipengaruhi oleh lingkungan dan sebaliknya, begitu pula proses keputusan dapat mempengaruhi lingkungannya.

Dalam keputusan bersifat tertutup, komputer bertindak sebagai sebuah alat penghitung yang dapat menghitung sebuah hasil hingga memiliki hasil optimum ataupun maksimal. Dalam keputusan bersifat terbuka, komputer bertindak sebagai alat bantu untuk kebutuhan pengambilan keputusan dalam masalah perhitungan, penyimpanan, pencarian kembali, kebutuhan menganalisa data dan sebagainya. Perancangan sistem tersebut memungkinkan manusia yang biasa mengambil keputusan mengalokasikan tugas bagi dirinya atau pada komputer.

## Klasifikasi

Klasifikasi merupakan suatu proses atau upaya untuk melakukan pemberian kategori terhadap suatu objek yang telah terdefinisikan sebelumnya berdasarkan model tertentu. *Data* *mining* merupakan sebuah upaya untuk menjelaskan hasil dari analisa atas sekelompok data tentang masa lalu, sehingga memungkinkan untuk membuat prediksi terhadap masa depan. Proses klasifikasi umumnya dimulai dari didapatkannya sejumlah data yang kemudian menjadi sebuah acuan untuk membuat aturan klasifikasi data. Data yang didapatkan ini kemudian disebut dengan istilah *training sets*. Dari data *training sets* tersebut nantinya dapat menghasilkan sebuah model guna mengklasifikasikan data lainnya. Model tersebut lalu dipergunakan untuk menjadi acuan dalam upaya mengklasifikasikan data – data yang belum diketahui kelasnya atau belum terklasifikasi. Data tersebut biasa disebut dengan istilah *datasets set.* (Nugroho, Kusrini and Arief, 2015)*.*

## Algoritma Random Forest

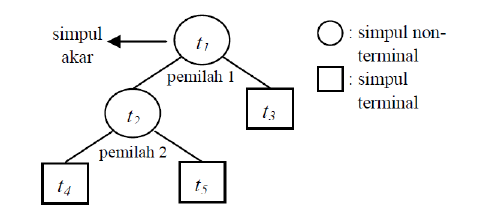
*Random Forest* diusulkan oleh Tin Kam Ho pada tahun 1995 dengan penelitiannya yang berjudul *Random Decision Forest* (Ho, 1995), yang kemudian dikembangkan Leo Breimann dan kemudian di patenkan pada tahun 2001 (Breiman, 2001). *Random Forest* merupakan salah satu metode *ensemble* yang memiliki tujuan untuk meningkatkan akurasi suatu klasifikasi data dari sebuah pemilah tunggal yang tidak stabil melalui kombinasi banyak pemilah dari suatu metode yang sama dengan proses *majority* voting untuk memperoleh prediksi pada klasifikasi akhir (van Wezel and Potharst, 2007)



**Gambar 2.1**: Contoh Random Forest

*Random Forest* adalah pengembangan lebih lanjut dari metode *Classification and Regression Tree* (CART) yang menerapkan metode *Bootstrap Aggregating* (*Bagging)* dan *Random Feature Selection*. Yaitu dengan melakukan modifikasi terhadap *bagging* yang membangun susunan pohon yang tidak berkorelasi lalu menghitung rata – ratanya. Dalam beberapa kasus, performa dari *Random Forest* mirip dengan *Boosting* namun lebih sederhana dan mudah untuk di *train* dan di *tune*. Sehingga, *Random Forests* menjadi salah satu metode klasifikasi populer dan di implementasikan di berbagai bidang (Hastie, Tibshirani and Friedman, 2001). CART sendiri merupakan metode klasifikasi data dengan cara melakukan eksplorasi terhadap data tersebut yang didasari pada teknik pohon keputusan. Pohon klasifikasi dihasilkan ketika variabel respon berupa data kategorikal, sedangkan pohon regresi dihasilkan saat variabel respon berupa data numerik. Untuk membangun pohon klasifikasi CART, meliputi tiga hal berikut yakni (Budi Adnyana, 2016):

1. Menentukan pemilah / pemisah *node* (*Splitting node*)
2. Penentuan simpul terminal
3. Penandaan label kelas.



**Gambar 2.2**: Struktur Pohon pada Metode CART (Budi Adnyana, 2016)

Dalam *Random Forest*, aka nada banyak pohon yang dihasilkan sehingga terbentuklah sebuah hutan (*forest*) dari kumpulan banyak pohon yang dihasilkan tersebut, kemudian dilakukan analisis terhadap kumpulan pohon tersebut. Pada kelompok data yang terdiri atas *n* amatan dan *p* peubah penjelas, algoritma *Random Forest* bekerja dengan cara:

1. Melakukan penarikan *sample /* data secara acak berukuran *n* dengan kemungkinan pengambilan data yang sama, tahap ini disebut tahapan *bootstrap*.
2. Dengan menggunakan contoh *bootstrap*. Pohon dibangun hingga mencapai ukuran maksimum (tanpa *pruning*). Pada setiap *node*, pemilihan *node* dilakukan dengan memilik *m* variabel penjelas secara acak, dimana *m* << *p*, lalu pemilah terbaik dipilih berdasarkan *m* variabel penjelas tersebut dimana tahapan ini disebut dengan tahapan *random feature selection*.
3. Ulang kembali langkah 1 dan 2 sebanyak *k* kali, sehingga terbentuklah sebuah hutan atas dari *k* pohon.

Untuk melakukan *splitting*, jika didalam *decision tree* menggunakan perhitungan terhadap *entropy* dan *gain*. Maka didalam metode *random forest* menggunakan *Gini index.* Yang mana sebelumnya juga dilakukan *feature selection*, yakni perhitungan untuk menentukan *feature* mana yang menjadi aturan dari *node* pohon keputusan. Untuk perhitungan *Gini index* terhadap *classification*, dilakukan perhitungan sebagai berikut (Breiman and Cutler, 2005):

|  |
| --- |
| Dimana: merupakan proporsi dari kelas *k* yang ada di node kiri dan  merupakan proporsi dari kelas *k* yang ada di node kanan. |

Untuk mendapatkan hasil yang optimal, metode *random forest* harus menentukan *m* jumlah variabel penjelas (*feature)* yang diambil secara acak dan juga untuk menentukan berapa banyak *k* pohon yang akan dibentuk. Nilai *k* yang disarankan untuk digunakan pada metode *bagging* yang telah di uji cobakan adalah *k* = 50 dimana nilai *k* tersebut memberikan hasil yang memuaskan untuk klasifikasi (Breiman, 2001). Sedangkan apabila nilai *k* yang diberikan adalah ≥100, nilai tersebut cenderung menghasilkan tingkat misklasifikasi yang rendah (Sutton, 2004).

Hasil respons dari suatu observasi diprediksi dengan cara menggabungkan (*Aggregating*) hasil prediksi *k* pohon. Dalam hal melakukan klasifikasi, hasil akhir dari suatu klasifikasi yang terbentuk dari beberapa pohon tersebut kemudian dipilih berdasarkan *Majority* *Vote* (suara terbanyak). Untuk menduga adanya kesalahan dalam klasifikasi *random forest*, dapat dilakukan dengan perolehan *error* OOB. Data OOB atau *Out of bag* merupakan data yang tidak termuat ketika melakukan pembuatan *bootstrap*. Adapun perolehan error OOB dapat dilakukan dengan cara sebagai berikut (Budi Adnyana, 2016):

1. Melakukan prediksi terhadap setiap data OOB pada pohon yang terbentuk.
2. Secara rata – rata, setiap observasi dari kelompok data asli akan menjadi data OOB sekitar 36% dari banyak pohon yang telah dihasilkan. Untuk itu, pada langkah 1, masing – masing observasi terhadap kelompok data yang asli mengalami sebanyak sekitar sepertiga kali dari banyak pohon. Jika *a* adalah sebuah pengamat dari kelompok data asli, maka hasil prediksi pada *random forest* terhadap *a* adalah gabungan dari hasil prediksi setiap kali *a* menjadi data OOB.
3. Error OOB dihitung dari proporsi data yang telah menjadi misklasifikasi dari hasil prediksi *random forest* dari seluruh observasi kelompok data yang asli.

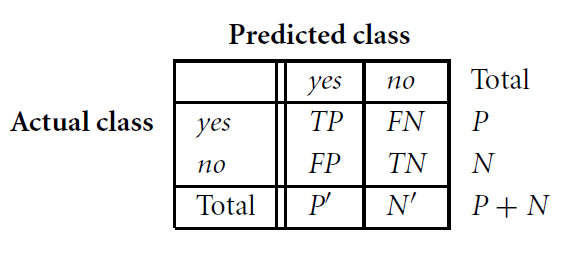
Sangat disarankan untuk melakukan pengamatan lebih teliti terhadap error OOB ketika *k* kecil, dan kemudian memilih *m* yang menghasilkan error OOB paling kecil. Apabila *random* *forest* akan dilakukan dengan menggunakan *variable importance*, maka disarankan untuk menggunakan banyak pohon, misalnya 1000 pohon atau lebih. Jika variabel penjelas yang dianalisis sangat banyak, nilai tersebut dapat menjadi lebih besar agar *variable importance* yang akan dihasilkan semakin stabil (Breiman and Cutler, 2005).

Dengan diterapkannya metode *random forest* pada permasalahan untuk rekomendasi kelayakan kredit, diharapkan dapat membantu pihak Bank dalam meninjau calon pemohon kredit untuk menekan dan mengurangi adanya *bad loan* di Bank BRI kantor cabang pelaihari untuk periode berikutnya.

## *Confusion Matrix* dan *Accuracy*

*Confusion matrix* merupakan salah satu alat pengukur yang memiliki fungsi untuk melakukan pengukuran ketika melakukan analisis terhadap sebuah *classifier*. Apakah *classifier* tersebut baik dalam hal mengenali *tuple* dari kelas yang berbeda atau tidak. Ketika *classifier* melakukan klasifikasi dan memiliki data yang bernilai benar, maka nantinya nilai dari *True-Positive* dan *True-Negative* berperan dalam memberikan informasi tersebut. Sedangkan apabila *classifier* memiliki kesalahan ketika melakukan klasifikasi data, maka nilai dari *False-Positive* dan *False-Negative* akan memberikan informasi tersebut (Han, Kamber and Pei, 2011).

Adapun bentuk *confusion matrix* serta penjelasannya dapat dilihat pada gambar 2.3:



**Gambar 2.3**: Confusion Matrix menampilkan total positive dan negative tuple (Han, Kamber and Pei, 2011).

1. TP (*True Positive*), merupakan jumlah data dengan nilai sebenarnya positif dan nilai prediksi positif.
2. FP (*False Positive*), merupakan jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif.
3. FN (*False Negative*), merupakan jumlah data dengan nilai sebenarnya positif dan nilai prediksi negatif.
4. TN (*True Positive*), merupakan jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif.

*Confusion Matrix* dapat digunakan untuk mengukur sebuah evaluasi, misalnya mengukur *accuracy. Accuracy* yang dimaksudkan adalah presentase keakuratan dari set tes *tuple* yang diklasifikasikan dengan benar oleh *classifier*. Adapun rumusnya sebagai berikut:

Perhitungan untuk *accuracy* juga dapat digunakan dengan menambahkan pengukuran *sensitivity* dan *specificity*. Dimana *sensitivity* juga di anggap sebagai *True-Positive* (proporsi dari *tuple* positif yang telah di identifikasi dengan benar)dan *specificity* dianggap sebagai *True-Negative* (proporsi dari *tuple* negatif yang telah di identifikasi dengan benar). Rumus dari *sensitivity, specificity* dan *accuracy* dengan menambahkan dua pengukuran tersebut adalah sebagai berikut:

## Analisa Kredit

Dalam upaya untuk memberikan resiko terkecil dalam memberikan pinjaman (kredit) dalam hal ini juga menghitung resiko kredit. Pihak bank biasanya terlebih dahulu melakukan pertimbangan terhadap beberapa hal terkaid dengan itikad baik untuk membayar kembali (*willingness to pay*) dan kemampuan untuk membayar (*ability to pay*) dari calon nasabah untuk melakukan pelunasan kembali pinjaman beserta bunga yang telah disepakati sebelumnya. Hal – hal tersebut dijelaskan dalam sebuah istilah yang apabila didalam dunia perbankan disebut “*The Five C of Credit Analysis*”, sebagai berikut (Pandie, 2012):

1. *Character* atau aspek karakter, berisi tentang data akan penilaian terhadap karakter dari calon debitur yang dilihat dari berbagai sifat. Misalnya watak, kemauan, kejujuran dan pengalaman hutang masa lalu.
2. *Capacity* atau aspek kemampuan, berisi tentang data akan penilaian terhadap kemampuan berusaha, kemampuan pasaaran, kemampuan membayar kembali hutangnya masa lalu dan hubungan dengan rekan usahanya.
3. *Capital* atau aspek permodalan, berisi tentang data akan penilaian terhadap data keuangan dari calon debitur yang meliputi harta lancer, harta tetap, hutang dan sebagainya.
4. *Conditions of Economy* atau aspek kondisi ekonomi, berisi tentang data akan penilaian terhadap kondisi usaha, kondisi rumah tangga, kondisi usaha yang berkaitan dengan kondisi ekonomi pada umumnya, dan sebagainya.
5. *Collateral* atau aspek jaminan, berisi tentang data akan penilaian terhadap agunan yang akan disediakan kebendaannya, keberadaannya, kondisi jaminannya, nilai jualnya, penilaian terhadap kelayakannya, dan sebagainya.

## Database

*Database* apabiladi definisikan secara umum adalah koleksi dari data – data yang terorganisasi sedemikian rupa sehingga dapat dengan mudah disimpan dan dimanipulasi (diperbarui, dicari, diolah dengan perhitungan tertentu, serta dihapus). Data merupakan fakta – fakta tentang segala sesuatu yang dapat direkan dan disimpan pada media komputer. Pada saat ini, data tidak hanya mengandung text saja, melainkan objek seperti dokumen, gambar hingga suara (Qomary, 2018).

## Integrasi Keilmuan

Kredit merupakan salah satu jasa yang disediakan oleh bank. Kredit merupakan istilah lain dari hutang. Bedanya, pengajuan kredit kepada bank dapat dilakukan dengan berbagai cara, ada yang mengajukan kredit dengan melakukan akad terlebih dahulu, mengajukan kredit dengan mengajukan jaminan (menggadaikan). Dan yang jelas, hutang piutang harus dicatat dan nantinya harus di bayarkan kembali seperti perjanjian awal.

Sebagaimana yang tertulis dalam Al-Qur’an surah Al-Baqarah ayat 282 yang memiliki arti sebagai berikut:

“*Hai orang-orang yang beriman, apabila kamu bermu'amalah tidak secara tunai untuk waktu yang ditentukan, hendaklah kamu menuliskannya. Dan hendaklah seorang penulis di antara kamu menuliskannya dengan benar. Dan janganlah penulis enggan menuliskannya sebagaimana Allah mengajarkannya, meka hendaklah ia menulis, dan hendaklah orang yang berhutang itu mengimlakkan (apa yang akan ditulis itu), dan hendaklah ia bertakwa kepada Allah Tuhannya, dan janganlah ia mengurangi sedikitpun daripada hutangnya. Jika yang berhutang itu orang yang lemah akalnya atau lemah (keadaannya) atau dia sendiri tidak mampu mengimlakkan, maka hendaklah walinya mengimlakkan dengan jujur. Dan persaksikanlah dengan dua orang saksi dari orang-orang lelaki (di antaramu). Jika tak ada dua oang lelaki, maka (boleh) seorang lelaki dan dua orang perempuan dari saksi-saksi yang kamu ridhai, supaya jika seorang lupa maka yang seorang mengingatkannya. Janganlah saksi-saksi itu enggan (memberi keterangan) apabila mereka dipanggil; dan janganlah kamu jemu menulis hutang itu, baik kecil maupun besar sampai batas waktu membayarnya. Yang demikian itu, lebih adil di sisi Allah dan lebih menguatkan persaksian dan lebih dekat kepada tidak (menimbulkan) keraguanmu. (Tulislah mu'amalahmu itu), kecuali jika mu'amalah itu perdagangan tunai yang kamu jalankan di antara kamu, maka tidak ada dosa bagi kamu, (jika) kamu tidak menulisnya. Dan persaksikanlah apabila kamu berjual beli; dan janganlah penulis dan saksi saling sulit menyulitkan. Jika kamu lakukan (yang demikian), maka sesungguhnya hal itu adalah suatu kefasikan pada dirimu. Dan bertakwalah kepada Allah; Allah mengajarmu; dan Allah Maha Mengetahui segala sesuatu*.”

Dari ayat di atas dijelaskan bahwa sebagai manusia dan ingin melakukan utang piutang dengan seseorang, hendaknya dituliskan segala utang piutangnya baik yang kecil maupun yang besar. Tak lupa menuliskan pula jumlah, dan tempo pembayarannya agar menghindari ketidak adilan dan keraguan. Selanjutnya pula dijelaskan hendaknya adanya seorang juru tulis pula yang menuliskan perihal utang piutang tersebut dengan tidak merusak sedikitpun dari perjanjian dan jumlah utang yang telah dikatakannya. Juru tulis tersebut adalah orang adil yang tidak memihak sebelah pihak. Hendaknya pemberi utang dan penerima utang mengutarakan maksudnya satu sama lain agar ditulis oleh juru tulis dan tidak mengurangi sedikitpun hak orang lain demi kepentingan pribadi (Al-Maraghi, 1986).

Adapun maksud dari tafsir ayat di atas apabila di integrasikan dengan penelitian ini adalah pencatatan utang yang dapat disamakan dengan *database*. *Database* pada sistem rekomendasi kelayakan kredit ini di asumsikan mencatat data dari para debitur yang mengajukan pinjaman yang nantinya akan diproses terlebih dahulu oleh sistem lalu disetujui oleh pihak atasan. Data debitur akan tersimpan terlebih dahulu didalam *database* sehingga dapat di simpulkan bahwa penyimpanan data calon debitur pada *database* tersebut merupakan salah satu upaya pencatatan utang piutang yang telah di jelaskan pada ayat diatas.

Adapula ayat yang membahas tentang segera membereskan hutang. Yang dimaksudkan adalah apabila telah diikat perjanjian hutang untuk jangka waktu tertentu, maka wajiblah janji itu ditepati dan pihak yang berhutang perlu membereskan hutangnya menurut perjanjian tersebut (Ya’qub, 1992). Menepati janji merupakan hal yang wajib dan setiap orang sangat dipertanggung jawabkan akan janji-janjinya. Sebagaimana firman Allah yang tertulis dalam Al-Qur’an surah Al-Isra potongan ayat 34 yang memiliki arti:

“*. . . . Dan penuhilah janji, sesungguhnya janji itu pasti dimintai pertanggungjawabannya.”*

Ayat tersebut apabila di integrasikan ke-ilmuannya dengan penelitian ini, penulis mencoba untuk membangun sebuah sistem pendukung keputusan yang dapat memberikan keputusan apakah seseorang layak dan mampu mengemban tanggung jawab apabila diberikan kredit atau utang berdasarkan kemampuan jasmani dan rohani seseorang tersebut.

Diasumsikan bahwa dengan adanya penelitian ini nantinya dapat dilakukan pencatatan utang piutang oleh seseorang dalam hal ini pencatatan data calon debitur yang mengajukan pinjaman kedalam *database*. Serta menemukan prediksi untuk keputusan terhadap sebuah pinjaman dari calon debitur yang bertanggung jawab.

# BAB III METODOLOGI PENELITIAN

Sistem rekomendasi kelayakan kredit ini adalah sistem yang menjadi sebuah alat bantu untuk membantu pihak bank mendapatkan kemudahan ketika peninjauan calon nasabah atau debitur yang sedang melakukan pengajuan, apakah nantinya calon nasabah atau debitur tersebut dapat dikategorikan menjadi *good loan* atau *bad loan*.

Sistem rekomendasi kelayakan ini berisi beberapa parameter yang telah ditentukan yang sebelumnya dianalisis terlebih dahulu dalam analisis kredit 5C. sehingga bertujuan untuk meminimalisir pihak bank mendapati *bad loan* dan meningkatkan *good loan*.

## Desain Penelitian

Adapun desain penelitian dideskripsikan menggunakan alur *flowchart*. Hal ini dilakukan oleh peneliti bertujuan agar mempermudah mencerna informasi dari alur penelitian, berikut alur dari penelitian ini:



**Gambar 3.1**: Alur Penelitian

Berikut adalah pembahasan dari proses alur metodologi yang telah tergambarkan diatas:

## Perumusan Masalah

Pada tahap ini, penulis merumuskan masalah yang telah menjadi latar belakang untuk membuat penelitian ini. Masalah yang di angkat pada penelitian ini seperti halnya tertera dalam latar belakang yakni mengenai rekomendasi kelayakan kredit. Rekomendasi kelayakan kredit merupakan salah satu upaya dari pihak bank untuk menghitung resiko kredit dari calon nasabah atau debitur baru. Dengan melakukan metode klasifikasi data mining dari data kredit yang sudah ada, sehingga setelah melakukan perhitungan pada data pengajuan yang baru didapatkan sebuah hasil apakah nanti pengajuan tersebut akan masuk *good loan* atau *bad loan*. Yang nantinya keputusan lebih lanjut masih berada di tangan pihak yang terkait dalam hal ini manajer.

## Studi Pustaka

Adapun dalam tahapan ini penulis melakukan studi pustaka yakni melakukan pendekatan lebih lanjut terhadap klasifikasi dalam data mining. Didapatkan hasil untuk melakukan klasifikasi dalam data mining ada beberapa metode yang populer diantara nya *KNN*, *Random Forest*, *Logistic Regression*, *Decision Tree*, dll. Setelah melakukan perbandingan terhadap studi pustaka lebih lanjut, didapatkan bahwa *Random Forest* merupakan metode yang populer dan memiliki akurasi prediksi yang paling tinggi dibandingkan metode klasifikasi lainnya dalam data mining. Sehingga, dalam penelitian ini penulis menggunakan metode *Random Forest* untuk proses klasifikasi.

## Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini dengan menggunakan metode – metode yang diharapkan dapat mengumpulkan data secara terperinci dan jelas, metode yang digunakan adalah:

1. Melakukan wawancara kepada pihak Bank BRI Kantor Cabang Pelaihari, dalam hal ini melakukan dengan Asisten Manajer untuk mengetahui permasalahan apa saja yang harus diperhatikan dalam proses penentuan kelayakan kredit pada calon debitur, memperoleh deskripsi kriteria untuk penentuan kelayakan kredit, memperoleh data mengenai kategori calon debitur.
2. Melakukan studi literatur lebih lanjut dalam memahami metode *Random Forest* dalam proses penentuan kelayakan kredit calon debitur baru berdasarkan kriteria yang ada.
3. Melakukan penghitungan terhadap data sekunder. Data sekunder yang dimaksud adalah data yang telah didapatkan dari *database* Bank BRI Kantor Cabang Pelaihari.

Adapun data sekunder yang dapat di ambil dari Bank BRI Kantor Cabang Pelaihari memiliki atribut dan di deskripsikan sebagai berikut:

**Tabel 3.1**: Deskripsi Atribut Data

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Nama Data** | **Jenis Data** | **Deskripsi** |
| 1. | Usia | Kuantitatif | Digunakan untuk klasifikasi dalam perihal segmentasi usia.  Contoh: ‘>60’, ‘<60’, dan ‘≤50’ |
| 2. | Alamat | Kualitatif | Digunakan untuk klasifikasi dalam perihal segmentasi daerah asal dari debitur.  Contoh: ‘Bati – Bati’, ‘Sungai Kiram’, ‘Pelaihari’ |
| 3. | Jenis Pekerjaan | Kualitatif | Digunakan untuk klasifikasi dalam perihal segmentasi jenis pekerjaan debitur.  Contoh: ‘Swasta’, ‘Wiraswasta’ |
| 4. | Penghasilan Perbulan | Kuantitatif | Digunakan untuk klasifikasi dalam perihal penghasilan dari debitur.  Contoh:‘>15.000.000’,‘≤30.000.000’ |
| 5. | Status Pernikahan | Kualitatif | Digunakan untuk klasifikasi dalam perihal segmentasi status pernikahan dari debitur.  Contoh: ‘Menikah’, ‘Belum Menikah’ |
| 6. | Jumlah Tanggungan | Kuantitatif | Digunakan untuk klasifikasi dalam perihal berapa jumlah kepala yang ditanggung oleh debitur.  Contoh: ‘4’, ‘3’, ‘2’ |
| 7. | Status BI Checking | Kualitatif | Digunakan untuk klasifikasi dalam perihal status BI Checking debitur.  Contoh: ‘Normal’, ‘Macet’, ‘Lancar’ |
| 8. | Status Sebagai Nasabah | Kualitatif | Digunakan untuk klasifikasi dalam perihal status apakah debitur sudah menjadi nasabah BRI atau belum.  Contoh: ‘Aktif’, ‘Tidak Aktif’ |
| 9. | Tujuan Kredit / Jenis Kredit / Sektor Kredit | Kuantitatif | Digunakan untuk klasifikasi dalam perihal tujuan dari kredit atau pada sektor apa kredit tersebut diajukan oleh debitur.  Contoh: ‘Tambang’, ‘Usaha’, ‘KUR’ |
| 10. | Jumlah Kredit Yang Di Ajukan | Kuantitatif | Digunakan untuk klasifikasi dalam perihal jumlah kredit yang akan di ajukan debitur.  Contoh:‘>100.000.000’,‘≤100.000.000’ |
| 11. | Jangka Waktu Kredit | Kuantitatif | Digunakan untuk klasifikasi dalam perihal jangka waktu atau tenor yang akan diajukan debitur.  Contoh: ’12 bulan’, ’36 bulan’, ’48 bulan’ |
| 12. | Jaminan | Kualitatif | Digunakan untuk klasifikasi dalam perihal bentuk jaminan yang di ajukan oleh debitur.  Contoh: ‘Sertifikat Rumah’, ‘Sertifikat Tanah’ |
| 13. | Status Pinjaman | Kualitatif | Digunakan untuk klasifikasi dalam perihal status terkini dari pinjaman debitur.  Contoh: ‘Lancar’, ‘Macet’ |

## Pengolahan Data

Dengan data yang berhasil dikumpulkan lalu diproses melalui tahap pengolahan data (*processing data*). Adapun tahapan - tahapan dalam upaya pengolahan data yakni sebagai berikut:

1. Hasil studi literatur akan digunakan sebagai dasar acuan dalam proses pembuatan model prediksi yang dihasilkan dari perhitungan metode *Random Forest* dalam membuat keputusan penentuan kelayakan kredit kepada calon debitur.
2. Melakukan analisa dari data yang diperoleh melalui wawancara mengenai proses bisnis dan permasalahan dalam penentuan kelayakan kredit kepada calon debitur.
3. Melakukan analisa dari data yang diperoleh melalui wawancara mengenai deskripsi kriteria yang digunakan dalam penentuan kelayakan kredit kepada calon debitur.
4. Melakukan perhitungan dan analisa terhadap *history* data nasabah yang telah didapat dari bank dengan metode *Random Forest*. Yang digambarkan dalam alur sebagai berikut:



**Gambar 3.2**: Flowchart Pembuatan Pohon pada Metode Random Forests dalam tahap training

## Pengembangan Sistem

Pengembangan sistem untuk penentuan kelayakan kredit ini menggunakan teknik atau metode yang disebut dengan metode *waterfall*. Adapun dalam metode *waterfall* ini memiliki lima tahapan proses yang harus dilalui, yakni (Safitri and Supriyadi, 2015):

1. Analisis Kebutuhan
2. Desain Sistem
3. Implementasi
4. *Testing*
5. Perawatan

## Pengujian Sistem

Untuk pengujian sistem, sistem akan diuji dengan melakuka perbandingan hasil dari sistem yang telah dibangun dan dikembangkan dengan data *real* yang telah didapat. Lalu dengan menggunakan metode pengujian *confusion matrix* dan perhitungan *accuracy*, *precision*, *recall*.

* 1. **Tempat dan Waktu Penelitian**

Penelitian ini dilaksanakan dari April 2018 hingga Juli 2018. Lokasi yang digunakan dalam penelitian ini adalah Bank BRI Kantor Cabang Pelaihari.

Adapun alur waktu dalam penelitian ini dijabarkan dalam tabel *timeline* seperti berikut:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Kegiatan | April | | | | Mei | | | | Juni | | | | Juli | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Perumusan Masalah |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Studi Pustaka |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Pengumpulan Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Pengolahan Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Pengembangan Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Pengujian Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Kesimpulan dan Pelaporan |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# BAB IV HASIL DAN PEMBAHASAN

Berdasarkan dua rumusan masalah yang telah dibahas pada bab 1 terkait dengan pembuatan sistem rekomendasi kelayakan kredit serta melakukan evaluasi terhadap metode *Random Forest* maka pada bab ini penulis melakukan pemaparan hasil dan pembahasannya.

* 1. **Pembuatan model klasifikasi dengan metode *random forest***

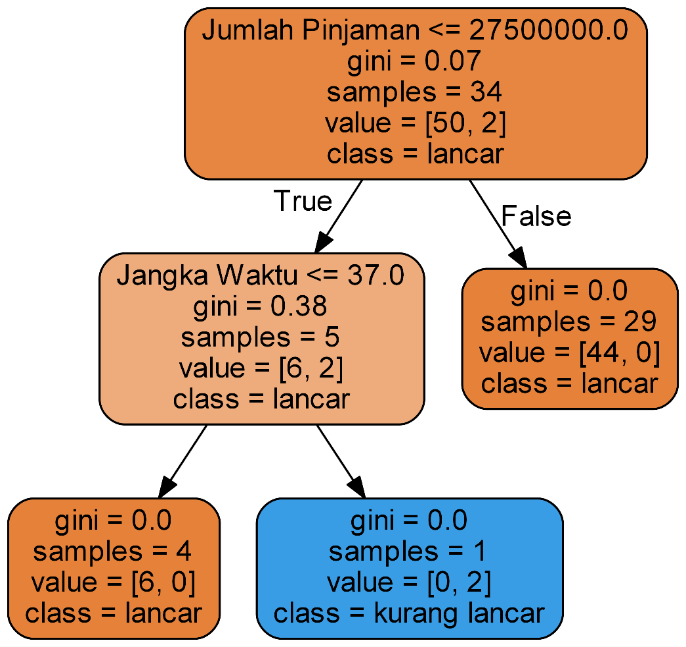
Untuk membuat sebuah model klasifikasi dengan menggunakan metode *random forest*. Ada beberapa tahap yang harus dilakukan dalam melakukan pembuatan model *machine learning*, terutama apabila ingin mengimplementasikan salah satu algoritma *machine learning* seperti *random forest*. Tahapan tersebut adalah *preprocessing data*, implementasi *random forest*, dan yang terakhir *predicting*.

## Pengenalan algoritma *Random Forest*

*Random forest* yang merupakan *ensemble method* atau improvisasi dari algoritma *decision tree* memiliki perbedaan yang sangat mendasar yaitu perhitungan untuk menentukan *node* dari *tree* nya. Apabila satu *decision tree* yang dibuat dalam algoritma *decision tree* menggunakan perhitungan *gain* dan *entropy*. Sedangkan *random forest* menggunakan perhitungan *gini* untuk menentukan *node* teratas hingga melakukan pemisahan *node* dari masing – masing *tree* yang akan terbuat.

Adapun perhitungan *gini impurity* mempengaruhi dari setiap *node* pemisah. Apabila perhitungan *gini* sudah memiliki hasil akhir = 0, maka perhitungan akan berhenti. Namun apabila perhitungan *gini* masih memiliki hasil akhir berupa angka, maka perhitungan akan tetap berlanjut.

Seperti yang dijelaskan pada bab sebelumnya, perhitungan *gini impurity* bergantung dari proporsi kiri dan proporsi kanan. Apabila data yang dimasukkan dalam perhitungan *gini* masih memiliki hasil akhir maka *node* akan memisah ke kiri, namun apabila perhitungan *gini* mencapai angka = 0 maka perhitungan akan berhenti. Penjelasan ini dapat dilihat dari gambar berikut:

****

**Gambar 4.1**: Salah satu tree yang terbentuk dari dataset pada algoritma random forest

Gambar 4.1 diatas menunjukkan salah satu dari beberapa tree yang dibuat dengan algoritma *random forest*. Tidak dapat dilacak *row* mana saja dari *dataset* yang dilakukan untuk pembuatan *tree* dari gambar tersebut. Ini dikarenakan algoritma *random forest* menggunakan *row* data secara acak untuk pembuatan masing – masing *tree* nya.

Untuk mencontohkan perhitungan *gini* dalam penentuan *node* teratas, maka akan terlebih dahulu perhatikan formula berikut:

|  |  |
| --- | --- |
|  | (4.1) |

Dimana *pi* merupakan probabilitas dari objek yang akan diklasifikasikan dalam kelas / *feature* tertentu. Selanjutnya *dataset* yang akan diuji cobakan untuk perhitungan *gini* akan disederhanakan terlebih dahulu demi kemudahan informasi untuk klasifikasi. Sehingga terbentuk tabel seperti berikut:

**Tabel 4.1**: Potongan *dataset* untuk ujicoba hitung *gini index* manual

|  |  |  |
| --- | --- | --- |
| Alamat | Pekerjaan | Status Kredit |
| Takisung | Karyawan Swasta | Macet |
| Pelaihari kota | Guru | Lancar |
| Bati Bati | PNS | Lancar |
| Pelaihari kota | Pensiunan | Lancar |
| Pelaihari kota | PNS | Lancar |
| Batu Ampar | Karyawan BUMN | Lancar |
| Pelaihari kota | Karyawan Swasta | Macet |
| Jorong | Karyawan Swasta | Macet |
| Pelaihari kota | Karyawan BUMN | Macet |
| Pelaihari kota | PNS | Lancar |

Jika diperhatikan pada tabel 4.1 diatas ada *feature* ’Alamat’, ‘Pekerjaan’ serta label ‘Status Kredit’ yang memiliki 10 *row* data, kemudian lakukan perhitungan *gini* dari masing – masing *feature*. Hitunglah terlebih dahulu masing – masing proporsi dari data diatas dalam hal ini *feature* ‘Alamat’. Sehingga apabila *feature* ‘Alamat’ dinotasikan kedalam tabel, proporsi data *feature* tersebut adalah sebagai berikut:

**Tabel 4.2**: Proporsi data dari *feature* ‘Alamat’

|  |  |
| --- | --- |
| Alamat | Proporsi Data |
| Takisung |  |
| Pelaihari Kota |  |
| Bati – Bati |  |
| Batu Ampar |  |
| Jorong |  |

Setelah didapatkan proporsi data dari *feature* ‘Alamat’, selanjutnya adalah mencari probabilitas *feature* ‘Alamat’ ketika data tersebut kemungkinan mendapatkan label dari ‘Status Kredit’. Apabila dinotasikan kedalam tabel maka akan terbentuk tabel sebagai berikut:

**Tabel 4.3**: Probabilitas Lancar dan Macet

|  |  |  |
| --- | --- | --- |
| Alamat | Probabilitas Lancar | Probabilitas Macet |
| Takisung |  |  |
| Pelaihari Kota |  |  |
| Bati – Bati |  |  |
| Batu Ampar |  |  |
| Jorong |  |  |

Setelah mendapatkan proporsi data dan probabilitas dari masing – masing data pada *feature* ‘Alamat’, selanjutnya adalah masukkan angka tersebut kedalam perhitungan *gini* yang dinotasikan sebagai berikut:

**Tabel 4.4**: Hasil perhitungan g*ini index*

|  |  |  |
| --- | --- | --- |
| Alamat | Gini Index Formula | Hasil Gini Index |
| Takisung |  | 0,88888889 |
| Pelaihari Kota |  | 0,22902494 |
| Bati – Bati |  | 0,97959184 |
| Batu Ampar |  | 0,97959184 |
| Jorong |  | 0,97959184 |

Dari tabel diatas, dapat dilihat hasil *gini index* dari masing – masing klasifikasi dari *feature* ‘Alamat’. Kemudian lakukan perhitungan untuk menentukan *gini impurity*. Seperti yang dinotasikan sebagai berikut:

|  |
| --- |
|  |

Lakukan perhitungan *gini index* diatas pada setiap *feature* yang ada untuk kebutuhan klasifikasi sehingga *feature* tersebut menghasilkan masing – masing *gini index* nya. Apabila perhitungan *gini index* juga dilakukan pada *feature* ‘Pekerjaan’, sehingga menghasilkan gini index sebagai berikut:

**Tabel 4.5**: Hasil *gini index* dari *feature* ‘Alamat’ dan ‘Pekerjaan’

|  |  |
| --- | --- |
| *Features* | *Gini Index* |
| Alamat | 0,52018141 |
| Pekerjaan | 0,73263889 |

Menurut *gini index* yang dihasilkan pada tabel diatas, *feature ‘*Alamat’ memiliki *gini* *index* terkecil. Sehingga, *feature* ‘Alamat’ menjadi *node* teratas dari salah satu *tree*. Untuk menentukan *splitting node*, kembali dilakukan perhitungan *gini index* dengan menjadikan *feature* ‘Alamat’ sebagai klasifikasi.

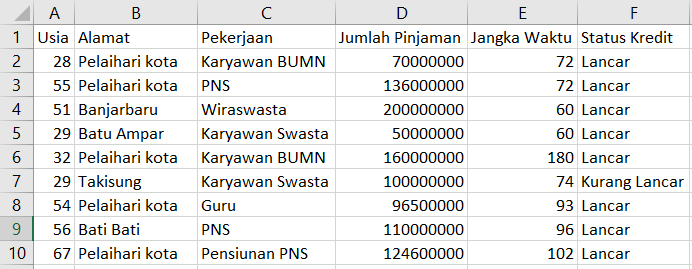
## *Preprocessing data*

Dalam melakukan *preprocessing data*, penulis mendapatkan *dataset* lalu melakukan *preprocessing data* secara manual untuk kebutuhan data agar statis. Pada tabel berikut adalah penjelasan dataset awal dan dataset yang telah dikonversi *preprocessing.*

**Tabel 4.6**: Bentuk Tipe Data Awal

|  |  |  |
| --- | --- | --- |
| **No** | **Nama Data** | **Tipe Data** |
| 1 | Usia | Integer |
| 2 | Alamat | Text |
| 3 | Pekerjaan | Text |
| 4 | Jumlah Pinjaman | Integer |
| 5 | Jangka Waktu | Integer |
| 6 | Status Kredit | Text |

Sebagai contoh, bentuk dari *dataset* yang dijelaskan pada tabel diatas dicontohkan dalam gambar sebagai berikut:



**Gambar 4.1**: Tabel berisi dataset sebelum dikonversi pada aplikasi Excel

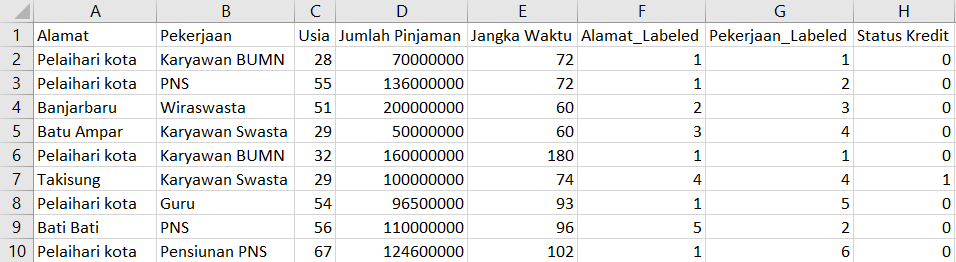
Setelah dilakukan *preprocessing data*, beberapa data diatas berubah tipe data nya menjadi sebagai berikut:

**Tabel 4.7:** Bentuk Tipe Data Akhir

|  |  |  |
| --- | --- | --- |
| No | Nama Data | Tipe Data |
| 1 | Usia | Integer |
| 2 | Alamat | Text |
| 3 | Pekerjaan | Text |
| 4 | Jumlah Pinjaman | Integer |
| 5 | Jangka Waktu | Integer |
| 6 | Alamat\_Labeled | Integer |
| 7 | Pekerjaan\_Labeled | Integer |
| 8 | Status Kredit | Integer |

Penambahan ‘Alamat\_Labeled’ dan ‘Pekerjaan\_Labeled’ pada kolom *dataset* dengan tipe data *integer* untuk pelabelan data atau kemudahan ketika *slicing data* ketika dilakukan *data mining*. Dikarenakan pengimplementasian *machine learning* dengan menggunakan algoritma *random forest* tidak dapat menggunakan tipe data *string* atau *text*. Sehingga dibuatlah kolom baru sebagai alias terhadap data asli. Lalu perubahan tipe data pada kolom ‘Status Kredit’ yang awalnya berupa *string* atau *text* yang berisikan ‘Lancar’ dan ‘Kurang Lancar’, tipe datanya berubah menjadi *integer* yang berisikan ‘0’ yang mendefinisikan ‘Lancar’ dan ‘1’ yang mendefinisikan ‘Kurang Lancar’. Pelabelan tersebut juga dilakukan untuk membuat data terbentuk menjadi sebuah kategori, sehingga dapat dengan mudah diklasifikasikan dengan sistem.

Sebagai contoh, bentuk perubahan *dataset* yang digambarkan pada tabel diatas dicontohkan dalam gambar sebagai berikut:



**Gambar 4.2**: Tabel berisi dataset setelah dikonversi pada aplikasi Excel

Setelah data selesai pada tahap *preprocessing*, dilanjutkan kepada tahap implementasi *random forest*.

## Implementasi *Random Forest*

Pengimplementasian algoritma *machine – learning­* *random forest* menggunakan bahasa *python* dengan bantuan *library* *sci-kit learn*. Adapun tahap – tahap pengimplementasian *machine learning* dengan menggunakan algoritma *random forest* untuk membuat *model* nya adalah sebagai berikut:

1. Menyiapkan *tools requirement*

Langkah pertama dalam pembuatan *model machine learning* dengan *algoritma random forest* adalah memastikan perangkat memiliki beberapa *tools requirement* seperti:

1. *Python* dengan versi terbaru
2. *Library* *Sci-kit Learn*
3. *Library Numpy*
4. *Library Pandas*
5. Anaconda, IDE untuk *python*
6. *Jupyter notebook*

Pada *jupyter notebook*, lakukan *import* terhadap *pandas* untuk kebutuhan manipulasi data dan *numpy* untuk kebutuhan perhitungan. Dengan menuliskan *syntax* pada *cell* yang tersedia secara berurutan. Contohnya seperti berikut:

|  |  |
| --- | --- |
| 1 | import pandas as pd  import numpy as np |

1. *Import Dataset*

Pastikan *dataset* yang telah dilakukan *preprocessing* sebelumnya format (.csv), dan lalu lakukan import dengan menuliskan *syntax* sebagai berikut:

|  |  |
| --- | --- |
| 2 | dataset = pd.read\_csv("converted.csv", ';') |

1. Memisahkan antara *features* dan *label*

Melakukan *data mining* tidak lepas dari kata *slicing*. Untuk itu dilakukan pemisahan antara *features* dan *label* yang masing – masing di notasikan sebagai X (*features*) dan y (*label*) dengan menuliskan *syntax* sebagai berikut:

|  |  |
| --- | --- |
| 3 | X = dataset.iloc[:, 2:7]  y = dataset.iloc[:,7] |

1. Melakukan pemisahan data untuk kebutuhan *Testing* dan *Training*.

*Testing* dan *Training* merupakan salah satu bentuk usaha pada *data mining* demi melakukan sebuah evaluasi pada suatu *model* yang akan dicoba untuk diterapkan. Berikut adalah *syntax* untuk melakukan *Testing* dan *Training*:

|  |  |
| --- | --- |
| 4 | from sklearn.model\_selection import train\_test\_split  X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.5, random\_state=0) |

1. Membuat *model* *machine learning*

Pada tahap ini akan mulai dilakukannya pembuatan *model* *machine learning* yang nantinya akan disimpan dan di ekstraksi untuk kebutuhan prediksi pada sistem rekomendasi kelayakan kredit. berikut adalah *syntax* untuk pembuatan *model machine learning*:

|  |  |
| --- | --- |
| 5 | from sklearn.ensemble import RandomForestClassifier  model=RandomForestClassifier(n\_jobs=2, random\_state = 0 )  model.fit(X\_train, y\_train)  model.predict(X\_test)  y\_pred = model.predict(X\_test) |

## Pengembangan sistem

Dalam pengembangan sistem pada penelitian ini digunakan metode *waterfall*. Adapun pada metode ini memiliki analisis kebutuhan, desain sistem, implementasi, *testing,* perawatan.

## Analisis kebutuhan sistem

1. Identifikasi masalah

Permasalahan yang terjadi pada penentuan kelayakan kredit pada Bank BRI Kantor Cabang Pelaihari adalah sebagai berikut:

1. Pengajuan kredit masih secara manual, nasabah terlebih dahulu harus pergi ke bank untuk proses pengajuan administrasi kredit sebelum nantinya ditindak lanjuti pengajuannya.
2. Belum adanya sistem rekomendasi yang memprediksi kelayakan kredit dari apakah calon nasabah yang akan mengajukan kredit.
3. Analisa sistem

Analisa sistem digunakan untuk mencari apa kebutuhan user yang terkait dan sesuai dengan sistem yang akan dibuat, berikut adalah analisa sistem:

1. Data calon nasabah untuk dilakukan perhitungan pada *machine* *learning* algoritma *random forest*. Dan juga untuk mencatat informasi nasabah tersebut.
2. Data prediksi calon nasabah dari hasil perhitungan algoritma *random forest* apakah nanti calon nasabah tersebut layak diberikan pinjaman atau tidak.
3. Kebutuhan data

Data yang dibutuhkan untuk mengembangkan sistem rekomendasi pada penelitian ini adalah sebagai berikut:

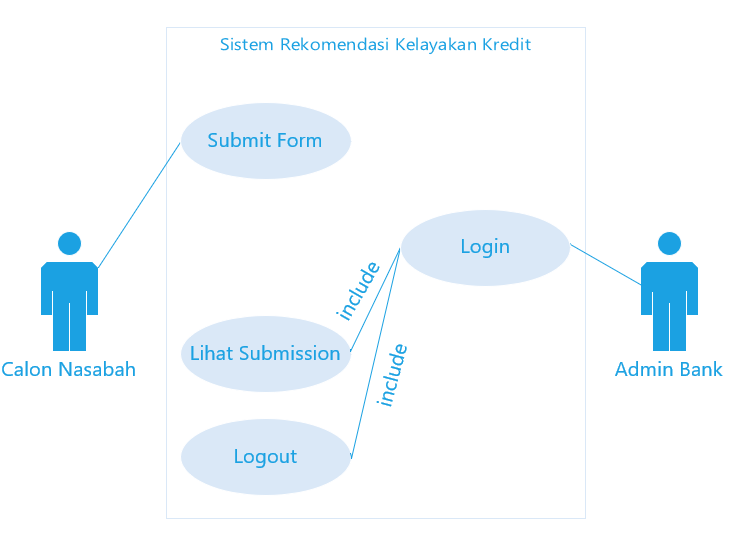
* 100 row record data kredit nasabah pada Bank BRI Kantor Cabang Pelaihari

## Desain sistem

Dalam desain sistem terdapat beberapa pembahasan diantaranya *use case diagram*, *activity diagram* dan desain *database*. Yakni sebagai berikut :

1. *Use case diagram*

*Use case* *diagram* berfungsi untuk menjelaskan peran – peran user (dalam hal ini disebut aktor) yang memiliki peran penting untuk mengorganisir dan memodelkan perilaku dari sistem yang akan dikembangkan. Dalam penelitian ini, pada *use case diagram* terdapat 2 aktor yang berperan dalam sistem rekomendasi kelayakan kredit yakni Admin Bank dan Calon Nasabah. Masing – masing aktor memiliki peran dan batasan – batasan tersendiri dalam sistem ini.



**Gambar 4.3**: Use Case Diagram pada Sistem Rekomendasi Kelayakan Kredit

Dapat dilihat bahwa pada *use case diagram* diatas ada 2 aktor yang dapat menjalankan sistem tersebut, yaitu Calon Nasabah dan Admin Bank. Aktor pertama yaitu, Calon Nasabah hanya dapat mengakses ‘Submit Form’. Calon Nasabah hanya dapat mengisi formulir pengajuan kredit, yang nantinya ketika *submit*, sistem akan melakukan penyimpanan data Calon Nasabah tersebut.

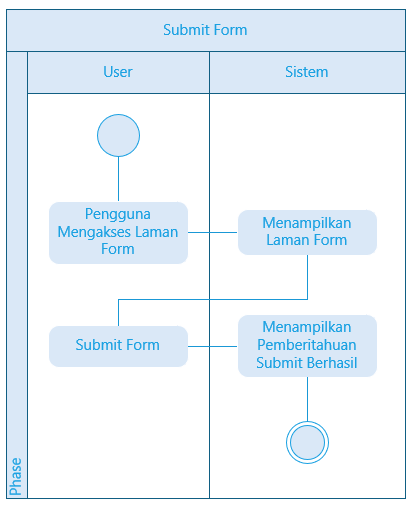
Aktor kedua adalah Admin Bank, dijelaskan pada diagram diatas Admin Bank dapat melakukan ‘*login’*, ‘lihat *submission’*, dan ‘*logout’*. *Login* kedalam sistem adalah akses atau wewenang yang hanya bisa dimiliki oleh Admin Bank. Karena untuk melakukan lihat *submission*, user harus terlebih dahulu *login*. Lihat *submission* merupakan laman yang berisi *record* *data* para Calon Nasabah yang telah berhasil melakukan *submit* pada ‘Submit Form’. Yang mana pada laman itu pula terlihat hasil prediksi dari Calon Nasabah, apakah dia layak untuk diberikan pinjaman atau tidak.

1. *Activity diagram*

*Activity diagram* adalah alur alur aktifitas yang akan di lakukan oleh user ketika menggunakan sistem. Diagram ini berfungsi untuk memperjelas dan mempermudah user dalam memahami alur sistem yang akan berjalan. Berikut adalah alur alur dari *activity diagram* pada sistem rekomendasi kelayakan kredit.

* 1. *Submit Form*

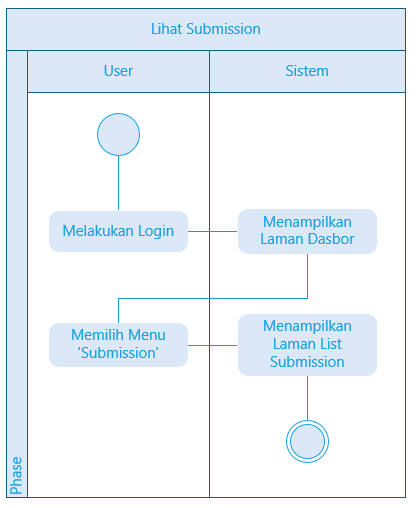
Pada *activity* *diagram* untuk ‘*Submit Form*’ adalah alur untuk memasukkan data Calon Nasabah kedalam sistem. Aktifitas ini hanya bisa dilakukan oleh *user* calon nasabah. Pengguna mengakses laman form terlebih dahulu, sehingga nantinya sistem menampilkan laman form. Lalu user mengisikan data lengkap pada form yang telah tersedia, dan melakukan submit. Sistem akan menampilkan pemberitahuan apabila form berhasil masuk kedalam sistem.



**Gambar 4.4**: Activity Diagram untuk submit form

* 1. Lihat *Submission*

Pada *activity* *diagram* untuk ‘Lihat *Submission*’ adalah alur untuk melihat hasil *input* dari *form* yang telah diisi dan di *submit* oleh calon nasabah. Aktifitas ini hanya bisa dilakukan oleh admin bank, yang mana untuk melakukannya pengguna harus melakukan *login* terlebih dahulu. Selanjutnya sistem akan menampilkan laman dasbor, pengguna selanjutnya memilih menu *submission* yang ada pada dasbor. Lalu nantinya sistem akan menampilkan laman *submission*, dan tersedia lah data – data calon nasabah beserta prediksinya. Pada alur ini, apabila pengguna ingin keluar dari sistem maka pengguna dapat melakukan *logout*.



**Gambar 4.5**: Activity Diagram untuk lihat submission

## Implementasi

Dalam tahap implementasi, akan dijelaskan sebuah pengembangan sistem rekomendasi kelayakan kredit hingga pada tahap pengimplementasiannya. Pengembangan sistem ini berbasis *webservice* dengan bahasa PHP. Berikut adalah beberapa *requirement* yang telah dilengkapi serta yang menjadi fokus utama pada penelitian kali ini:

1. PHP
2. HTML
3. *python*
4. *RestAPI*

Dikarenakan adanya banyak cara dalam memanipulasi tampilan dan data menggunakan PHP sehingga penulis memilih bahasa PHP untuk dasar pembuatan *webservice* dari sistem rekomendasi kelayakan kredit ini. *Python* digunakan sebagai bahasa untuk melakukan *prediksi* data yang telah didapat dari sistem dengan memasukkan data tersebut kedalam *model* perhitungan prediksi yang sudah terbuat sebelumnya. Untuk mengirim data yang didapatkan dari sistem, digunakan *restAPI* sebagai jembatan penghubung antara sistem dan *model* perhitungan prediksi. Hal ini dikarenakan perbedaan jenis bahasa yang digunakan, sistem dengan penggunaan *webservice* yang menggunakan bahasa HTML dan PHP, sedangkan *model* perhitungan prediksi dengan menggunakan bahasa *python*.

Selanjutnya akan dibahas tentang langkah – langkah secara spesifik tentang pembuatan model hingga implementasinya kedalam sistem. Yakni sebagai berikut:

1. Pembuatan file *main.py* dan *ml.py*

*File* *main.py* terbilang sangat penting karena memiliki peran sebagai *restAPI* yang akan mengirimkan data yang didapat dari sistem menuju *model* perhitungan prediksi. Sedangkan *model* perhitungan prediksi yang berisi algoritma *random forest* ada didalam *file ml.py*. Inilah alasan kenapa dua *file* ini penting dalam penelitian kali ini.

Pada sub bab diatas, dijelaskan bahwa kita telah membuat *model* perhitungan prediksi menggunakan algoritma *random forest*. Sehingga untuk itu dibuatlah fungsi untuk melakukan *train* data terhadap *model*, serta fungsi untuk melakukan prediksi pada data tertentu, dijelaskan seperti *source code* berikut:

|  |
| --- |
| def TrainData():    X = dataset.iloc[:, 2:7]  y = dataset.iloc[:,7]  X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.5, random\_state=0)  model = RandomForestClassifier(n\_jobs=2, random\_state = 0 )  model.fit(X\_train, y\_train)  return model  def PredictData(DataBaru):  model = TrainData()    inputDF = pd.DataFrame(data = DataBaru)  result = model.predict(inputDF)    return result |

Pada *source code* diatas fungsi TrainData() digunakan untuk pembuatan *model* dari algoritma *random forest* yang nantinya akan melakukan prediksi dengan menggunakan fungsi PredictData(). Sedangkan didalam fungsi PredictData(), ada bantuan fitur dari *pandas* untuk kebutuhan pembuatan *dataframe*. Karena untuk melakukan perhitungan terhadap *model*, data yang dimasukkan harus berupa *array* dalam bentuk *dataframe*.

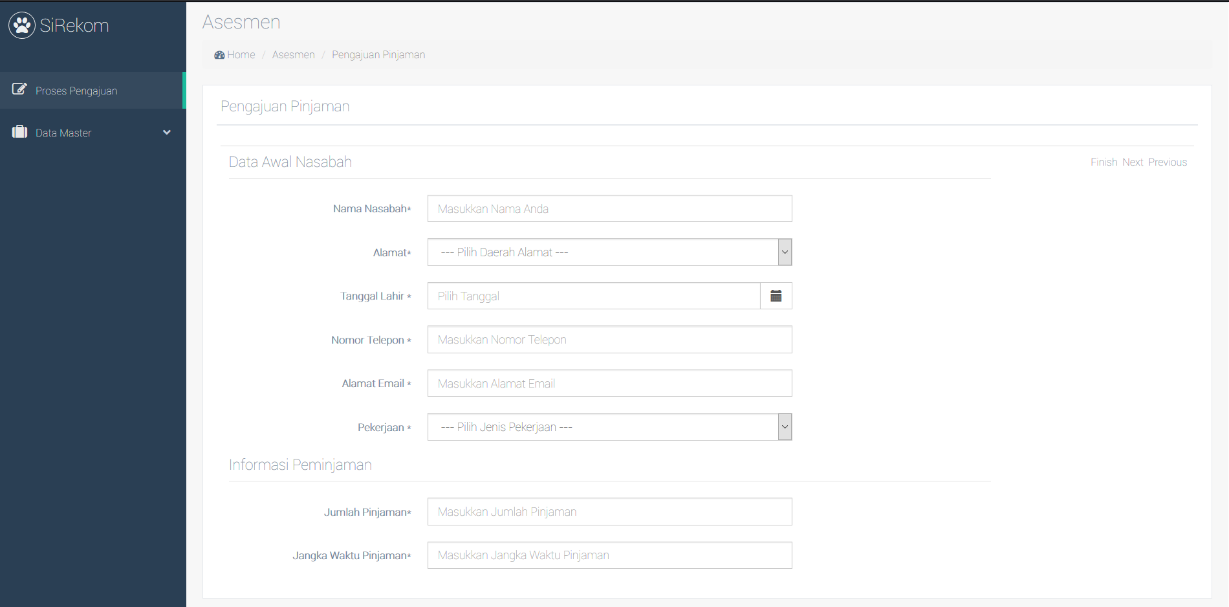
Berikutnya adalah *file main.py* dimana dilakukannya *restAPI*. Untuk kemudahan kebutuhan sistem digunakan *json* sebagai penotasi pertukaran data. Perhatikan *source code* berikut:

|  |
| --- |
| class MLService(Resource):  def post(self):  if not request.json:  return {  "error": True,  "message": "no data found"  }  dataInput = {  'Usia': [request.json["usia"]],  'Jumlah Pinjaman': [request.json["jumlah\_pinjaman"]],  'Jangka Waktu': [request.json["jangka\_waktu"]],  'Alamat\_Labeled': [request.json["alamat\_labeled"]],  'Pekerjaan\_Labeled': [request.json["pekerjaan\_labeled"]]  }  result = (ml.PredictData(dataInput))  hasilString = "Lancar"  if result[0] == 1:  hasilString = "Macet"    return {  "error": False,  "message": "Belajar Res",  "result": hasilString,  } |

Jika diperhatikan pada *source code* diatas, data yang akan dimasukkan kedalam sistem nantinya akan diambil oleh *json* yang telah diatur pada *syntax* dataInput. Kemudian ketika telah berhasil diambil, kemudian di arahkan ke *model* perhitungan prediksi pada *syntax* ml.PredictData().

1. Tampilan *form* pengajuan kredit

Berikut merupakan tampilan *form* pengajuan kredit pada sistem rekomendasi kelayakan kredit berbasis *webservice*:



Ada beberapa batasan dari sistem rekomendasi kelayakan kredit berbasis *webservice* ini. Diantaranya adalah pengisian *form* alamat serta pekerjaan yang menggunakan *dropdown*. Hal ini dikarenakan keterbatasan data yang dijadikan sebagai *train* untuk pemodelannya serta pengkategorian data pada *preprocessing* yang masih manual. Sehingga diasumsikan sistem hanya melakukan prediksi terhadap calon nasabah dengan *alamat* dan *pekerjaan* yang ada pada pilihan *dropdown* tersebut.

1. Tampilan *list submission*

Berikut merupakan tampilan *list submission* dari hasil *form* pengajuan kredit pada sistem rekomendasi kelayakan kredit berbasis *webservice*:

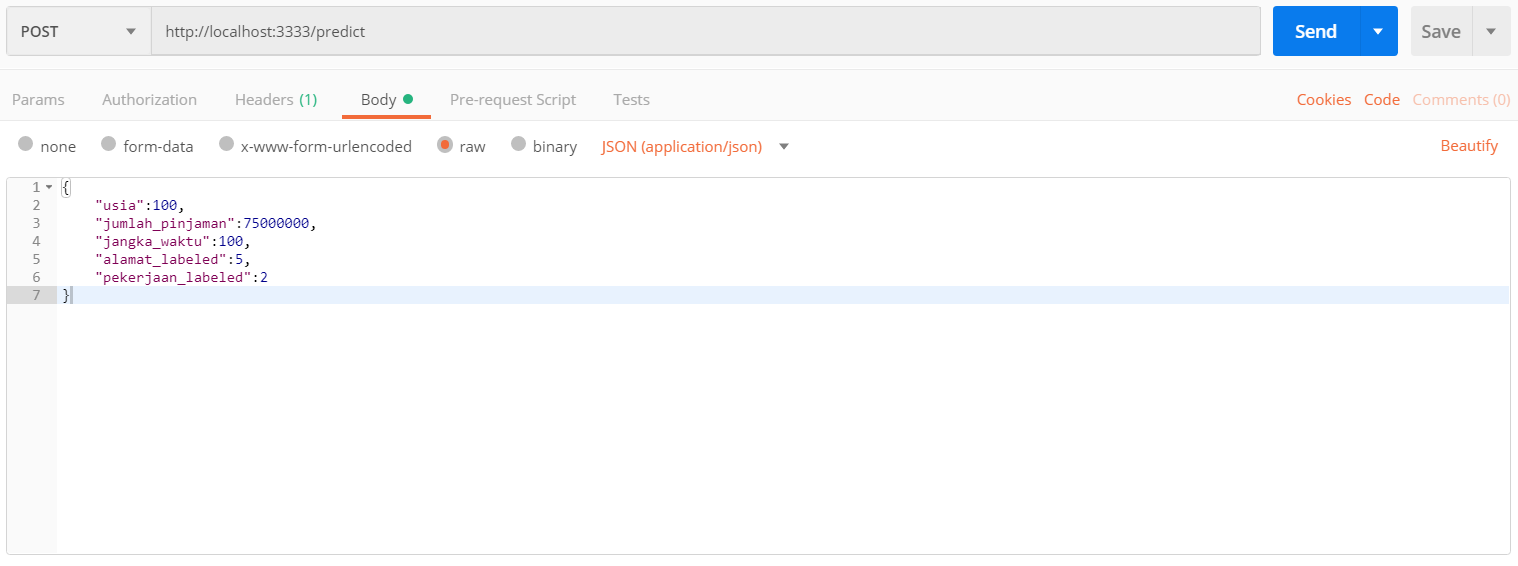
****

Untuk tampilan yang minimalis, diasumsikan bahwa data hasil prediksi yang ditampilkan pada tampilan *list submission* hanya ‘Nama’, ‘Email’, ‘Phone’ dan ‘Potensi Kredit’. Hal ini dilakukan agar pihak bank berfokus pada potensi kredit dari calon nasabah yang mengajukan, serta informasi kontak untuk menghubungi calon nasabah yang berpotesi tersebut.

## *Testing*

Berfokus terhadap *restAPI*, tahapan *testing* dilakukan untuk melakukan pengujian apakah *restAPI* yang dilakukan pada sistem telah berhasil. Untuk itu maka menggunakan bantuan pihak ketiga yaitu dukungan aplikasi *Postman*. *restAPI* yang diuji cobakan adalah *source code* dari file *main.py*.

Pada aplikasi *postman*, uji cobakan ‘POST’ pada *localhost* sesuai dimana aplikasi dibangun. Dalam penelitian ini, penggunaan untuk perhitungan prediksi berada pada url http://localhost:3333/predict. Dan kemudian pilih *body* > *raw* > *json* dan lalu isikan data seperti format *dictionary* dataInput pada *file main.py*, seperti gambar berikut:



Dan apabila ketika melakukan klik pada ‘send’ atau menekan tombol CTRL + Enter, maka apabila tidak terjadi *error* dan muncul *output* seperti gambar dibawah ini, maka berarti *restAPI* berhasil dilakukan.



## Perawatan

Pada tahap ini, sistem rekomendasi kelayakan kredit berarti sudah dalam keadaan berjalan, yang mana berarti ada banyak tambahan *record* data calon nasabah baru yang telah terprediksi. Maka dari itu untuk tahap perawatan berarti melakukan *train* ulang terhadap model *machine learning* algoritma *random forest* yang ada ketika ada cukup banyak data yang telah masuk. Sehingga diharapkan dengan semakin banyak data tersebut, makin kompleks model *machine learning* yang di hasilkan.

## Pengujian Sistem

Untuk melakukan pengujian sistem pada penelitian kali ini, penulis memilih metode pengujian untuk menghitung *accuracy*, *precision* dan *recall*. Hasil dari *data mining* dengan menggunakan algoritma *random forest* digambarkan dalam *confusion matrix* sebagai berikut:

|  |  |  |
| --- | --- | --- |
| Real  Prediksi | Kurang Lancar | Lancar |
| Kurang Lancar | 1 | 0 |
| Lancar | 5 | 98 |

Dikarenakan perubahan label dari ‘Lancar’ dan ‘Kurang Lancar’ menjadi ‘0’ dan ‘1’ pada tahap *preprocessing*, sehingga terjadi perubahan data terhadap ‘True’ dan ’False’ karena *library* dari scikit-*learn* menganggap bahwa ‘0’ dan ‘1’ adalah *binary code*. Sehingga untuk perbandingan, berikut adalah tabel asli dari *confusion matrix* (Kohavi and Provost, 1998):

|  |  |  |
| --- | --- | --- |
| Real  Sistem | *TRUE* | *FALSE* |
| *TRUE* | *True Positive (TP)* | *False Positive (FP)* |
| *FALSE* | *False Negative (FN)* | *True Negative (TP)* |

Sehingga, dari *confusion matrix* diatas didapatkan hasil *accuracy, precision* dan *recall* sebagai berikut:

|  |  |  |
| --- | --- | --- |
| *Accuracy* | = |  |
|  | = |  |
| *Precision* | = |  |
|  | = |  |
| *Recall* | = |  |
|  | = |  |

Untuk melakukan pengukuran terhadap klasifikasi kelayakan, dapat diukur dengan cara melihat skala pengukuran pada tabel berikut:

|  |  |  |
| --- | --- | --- |
| **No** | **Persentase Pencapaian** | **Klasifikasi**  **kelayakan** |
| 1 | 81% - 100% | Sangat Layak |
| 2 | 61% - 80% | Layak |
| 3 | 41% - 60% | Cukup Layak |
| 4 | 21% - 40% | Tidak Layak |
| 5 | 0% - 20% | Sangat Tidak Layak |

Menurut tabel diatas, tingkat *accuration* pada prediksi sistem rekomendasi kelayakan kredit mencapai 95,19%, tingkat *precision* mencapai 100%, akan tetapi tingkat *recall* hanya mencapai angka 16,67%. Untuk ukuran *accuration* dan *precision*, klasifikasi kelayakan menyatakan bahwa sistem ini ‘layak’ di implementasikan secara nyata. Namun tidak dengan hasil *recall* yang memiliki angka 16,67% yang mana dinyatakan “Sangat Tidak Layak”.

# BAB V PENUTUP

## 5.1. Kesimpulan

Penelitian yang memakan waktu selama kurang lebih dari empat bulan ini memiliki beberapa kesimpulan sebagai berikut:

1. Pengembangan sistem rekomendasi kelayakan kredit dengan menggunakan algoritma *random forest* sebagai *model* prediksi. Dan sistem ini dibangun dengan menggunakan pendekatan *waterfall*.
2. Pada pengujian *accuration*, *precision*, dan *recall* didapatkan hasil bahwa metode *random forest* dalam kasus penelitian ini memiliki masing – masing tingkat *accuration* sebesar 95,19%, *precision sebesar* 100%, serta 16,67%. Hal ini dikarenakan minimnya data ‘Kurang Lancar’ pada *dataset* yang dipelajari.

## Saran

Berdasarkan hasil penelitian ini, tentu masih terdapat banyak kekurangan yang dapat ditemui pada sistem rekomendasi kelayakan kredit ini. Maka dari itu, penulis memberikan beberapa saran guna memperbaiki dan meningkatkan kekurangan yang didapati dalam sistem rekomendasi kelayakan kredit ini, yakni:

1. Memperbanyak dataset serta memperpanjang *feature* dari dataset yang akan dipelajari, sehingga meningkatkan performa dari metode *random forest*.
2. Menggunakan studi kasus lain guna klasifikasi atau yang lainnya dengan menggunakan metode *random forest*.

# DAFTAR PUSTAKA

Afianto, M. F. *et al.* (2017) ‘Kategorisasi Teks pada Hadits Sahih Al-Bukhari menggunakan Random Forest’, 4(3), pp. 4874–4881.

Al-Maraghi, A. M. (1986) *Tarjamah Tafsir Al-Maraghi*.

Breiman, L. (2001) ‘Random forests’, pp. 1–33.

Breiman, L. and Cutler, A. (2005) *Random Forests*. Available at: https://www.stat.berkeley.edu/users/breiman/RandomForests/ (Accessed: 15 April 2019).

Budi Adnyana, I. M. (2016) ‘Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest (Studi Kasus : Stikom Bali)’, *CSRID (Computer Science Research and Its Development Journal)*, 8(3), pp. 201–208. doi: 10.22303/csrid.8.3.2016.201-208.

Han, J., Kamber, M. and Pei, J. (2011) *Data Mining Concept and Techniques*, *Data Mining: Concepts and Techniques*. doi: 10.1016/B978-0-12-381479-1.00001-0.

Hastie, T., Tibshirani, R. and Friedman, J. (2001) ‘The Elements of Statistical Learning The Elements of Statistical Learning’. Available at: https://web.stanford.edu/~hastie/Papers/ESLII.pdf.

Ho, T. K. (1995) ‘Random Decision Forest’, 47, pp. 4–5.

Kohavi, R. and Provost, F. (1998) ‘Glossary of Terms’, *Machine Learning*, 30, pp. 271–274. doi: 10.1016/B978-0-12-397026-8.00018-5.

Manurung, P. (2010) *Sistem Pendukung Keputusan Seleksi Penerima Beasiswa Dengan Metode AHP dan TOPSIS*. Universitas Sumatera Utara.

Nugroho, A., Kusrini, K. and Arief, M. R. (2015) ‘Sistem Pendukung Keputusan Kredit Usaha Rakyat PT. Bank Rakyat Indonesia Unit Kaliangkrik Magelang’, *Creative Information Technology Journal*, 2(1), p. 1. doi: 10.24076/citec.2014v2i1.33.

Nurdiyanto, H. and Minarto, S. Y. (2017) ‘Pada Koperasi Serba Usaha Berkah Tiram Jaya Menggunakan Metode Analytic Hierarchy Process ( Ahp )’.

Octabriyantiningtyas, D. (2016) *Sistem Pendukung Keputusan Kelayakan Pemberian Kredit Pada Bank Tabungan Negara (BTN) Menggunakan Algoritma C4.5*. Universitas Airlangga.

Pandie, E. S. Y. (2012) *Sistem Informasi Pengambilan Keputusan Pengajuan Kredit Dengan Algoritma K-Nearest Neighbour (Studi Kasus: Koperasi Simpan Pinjam)*. Universitas Diponegoro.

Putri, C. B. (2018) *Klasifikasi Nasabah Thera Bank Membeli Personal Loan Menggunakan Metode Klasifikasi Dalam Machine Learning Pendahuluan Metodologi Penelitian*. Institut Teknologi Sepuluh Nopember.

Qomary, L. N. (2018) *Rancang Bangun Sistem Pendukung Keputusan Pemilihan Program Latihan Fitnes Metode Case Based Reasoning*.

Safitri, S. T. and Supriyadi, D. (2015) ‘Rancang Bangun Sistem Informasi Praktek Kerja Lapangan Berbasis Web dengan Metode Waterfall’, pp. 3–8.

Sutton, C. D. (2004) ‘Classification and Regression Trees, Bagging, and Boosting’, *Handbook of Statistics*, 24(04), pp. 303–329. doi: 10.1016/S0169-7161(04)24011-1.

van Wezel, M. and Potharst, R. (2007) ‘Improved customer choice predictions using ensemble methods’, *European Journal of Operational Research*, 181(1), pp. 436–452. doi: 10.1016/j.ejor.2006.05.029.

Widiastuti, J. (2018) *Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode Random Forest Dengan Teknik Sampling Kelas Imbalanced*. Universitas Islam Indonesia.

Ya’qub, H. (1992) *Kode Etik Dagang Menurut Islam (Pola Pembinaan Hidup dalam Berekonomi)*.

Zhang, C. *et al.* (2017) ‘An up-to-date comparison of state-of-the-art classification algorithms’, *Expert Systems with Applications*. Elsevier Ltd, 82, pp. 128–150. doi: 10.1016/j.eswa.2017.04.003.