



Click here and write your Article Category

## Analisis Faktor Demografis dan Performansi Klasifikasi untuk Pengguna Kartu Kredit

Anisha Rizqy Aulya<sup>1</sup>, Aisyahna Nurul Mauliddina<sup>2</sup>, Arina Meitri Kurnia Larasati<sup>3</sup>, Faris Ahmad Saifuddin<sup>4</sup>

<sup>1</sup> Universitas Pertamina, Jakarta Selatan, 12220, Indonesia

<sup>2</sup> Universitas Pertamina, Jakarta Selatan, 12220, Indonesia

<sup>3</sup> Universitas Pertamina, Jakarta Selatan, 12220, Indonesia

<sup>4</sup> Universitas Pertamina, Jakarta Selatan, 12220, Indonesia

### ARTICLE INFORMATION

Received: February 00, 00

Revised: March 00, 00

Available online: April 00, 00

### KEYWORDS

Klasifikasi, *default payment*, *logistic regression*, *k nearest neighbor*, *random forest*, *support vector machine*

### CORRESPONDENCE

Phone: +62 (0751) 12345678

E-mail: [first\\_author@affiliation.xx.xx](mailto:first_author@affiliation.xx.xx)

### A B S T R A K

Pada dunia perbankan, sistem pembayaran nontunai menggunakan kartu kredit mampu memberikan keuntungan dan kemudahan untuk nasabah sebagai pemegang kartu kredit, maupun lembaga keuangan atau perbankan sebagai penyalur kartu kredit. Salah satu contoh nyatanya adalah pengguna kartu kredit aktif di Negara Taiwan mengalami peningkatan yang signifikan tiap tahunnya. Hal ini sejalan dengan peningkatan resiko, baik dalam segi nominal, maupun transaksi, karena adanya perilaku konsumtif yang mengakibatkan penambahan tagihan kartu kredit bagi nasabah. Dengan begitu, diperlukan adanya mitigasi resiko untuk melakukan evaluasi resiko bagi lembaga keuangan atau perbankan terkait dengan pemberian kredit. Teknik *data mining* klasifikasi dilakukan untuk menemukan model yang menjelaskan konsep dan karakteristik suatu kelas data. Pada klasifikasi, terdapat berbagai algoritma antara lain *K-nearest neighbor*, *Logistics regression*, *Random Forest*, dan *Support Vector Machine (SVM)*. Penelitian ini mengangkat permasalahan penentuan *default payment* pengguna kartu kredit berdasarkan pada data demografis dan riwayat pembayaran pelanggan. Tujuan penelitian ini adalah mengidentifikasi pengaruh faktor demografis terhadap tingkat *default payment* nasabah dan menentukan algoritma klasifikasi yang optimal untuk mengklasifikasikan data. Berdasarkan hasil pengolahan data dan analisis didapatkan hasil bahwa faktor demografis memiliki pengaruh terhadap *default payment* pengguna. Selain itu, algoritma *SVM* sebagai metode klasifikasi yang memiliki tingkat akurasi paling tinggi yaitu sebesar 0.99 atau sebesar 99% serta nilai *f1 score* yang paling tinggi dibandingkan dengan beberapa metode lainnya. Oleh karena itu, dapat disimpulkan bahwa algoritma *SVM* memiliki performansi yang lebih baik dibandingkan metode klasifikasi lainnya dalam menentukan *default payment* pengguna kartu kredit.

### INTRODUCTION

Fenomena penggunaan kartu kredit di dunia perbankan, merupakan salah satu wujud alat pembayaran nontunai. Sistem pembayaran nontunai yang mampu memenuhi ketersediaan dan tingkat layanan yang baik dapat mendukung proses stabilitas sistem keuangan dan mendukung kelancaran dari aktivitas perekonomian suatu negara (Waluyo, Dedy, & Agung, 2013).

Penggunaan kartu kredit termasuk dalam alat pembayaran bagi nasabah yang sudah dianggap memiliki tanggung jawab secara *financial*. Hal ini dikarenakan, penggunaan kartu kredit tanpa adanya jaminan dalam pembelian barang atau jasa. Proses penggunaan kartu kredit mendapat uang muka dari lembaga keuangan atau perbankan. Pihak lembaga

keuangan atau bank memberikan tagihan pinjaman ke nasabah atas transaksi melalui kartu kredit. Selain itu, terdapat bunga sebesar saldo pinjaman yang terdapat pada kartu kredit tersebut (Gunawan, 2017).

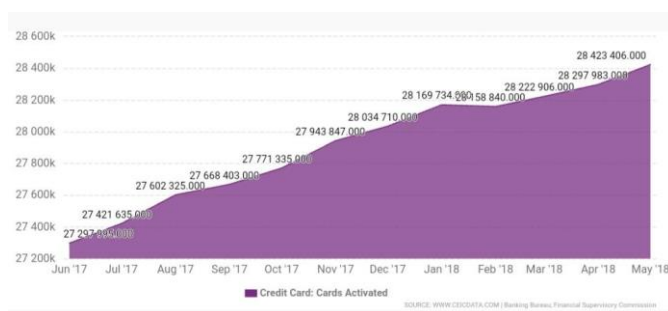
Fungsi dari penggunaan kartu kredit dapat digunakan sebagai alat pembayaran dan juga dapat sebagai sumber hutang atau kredit. Dengan adanya kartu kredit dapat memberikan keuntungan transaksi yang praktis, cepat, dan relatif aman. Selain itu, pemberian pinjaman uang muka dalam wujud saldo dapat digunakan secara tiba – tiba apabila pengguna membutuhkan pembayaran dalam keadaan mendesak. Berdasarkan frekuensi pengguna kartu kredit, terbagi atas pengguna aktif dan pengguna pasif. Pengguna aktif merupakan nasabah yang melakukan transaksi relatif sering menggunakan kartu kredit. Sedangkan, pengguna pasif merupakan nasabah yang jarang bahkan tidak

pernah menggunakan kartu kredit untuk bertransaksi. (Devlin, 2007 dalam (Gunawan, 2017)).

Pemberian layanan kartu kredit oleh pihak perbankan, bertujuan untuk mencapai visi dan misi yang dimiliki oleh pihak perbankan. Memperoleh keuntungan, mengembangkan usaha yang dimiliki nasabah, serta adanya peningkatan perekonomian suatu negara. Bagi nasabah yang memegang kartu kredit, beberapa hal yang dapat menjadi dampak positif dari penggunaan kartu kredit yaitu mempermudah transaksi di luar negeri, adanya catatan dari seluruh transaksi yang dilakukan, serta penggunaan kartu kredit bermanfaat untuk membantu dalam keadaan darurat.

Pada sisi perbankan, penggunaan kartu kredit sebagai pemberian sistem pembayaran nontunai memiliki resiko yang tinggi, baik berdasarkan sisi nominal maupun sisi transaksi. Hal ini sejalan dengan cara nasabah menggunakan kartu kredit. Keberadaan kartu kredit, selain memberikan dampak positif bagi nasabah, penggunaan kartu kredit juga dapat memberikan perilaku konsumtif yang mengakibatkan penambahan tagihan kartu kredit. Dengan begitu, diperlukan adanya evaluasi tingkat resiko bagi lembaga keuangan atau perbankan terkait dengan pemberian kredit pada nasabah yang memenuhi syarat (Yunindya, Kudus, & Yanti, 2016).

Penelitian terkait dengan permasalahan penggunaan kartu kredit telah dilakukan. Penggunaan kartu kredit aktif di Negara Taiwan menunjukkan adanya peningkatan dari beberapa tahun.



**Gambar 1** Data Pengguna Kartu Kredit Taiwan  
Credit Cards: Activated;  
Sumber: www.ceicadara.com

Pengguna kartu kredit di Taiwan pada Mei 2018 tercatat sebanyak 28.423.406 unit. Terjadi peningkatan dari jumlah pengguna kartu kredit pada bulan sebelumnya yaitu 28.297.983 unit untuk April 2018. Rata – rata kartu kredit yang diaktifkan setiap bulan dari Juni 2004 hingga Mei 2018 sebesar 21.916.623 unit. Data kartu kredit yang berstatus aktif di CEIC akan dilaporkan oleh *Banking Bureau, Financial Supervisory Commission*. Data tersebut masuk ke dalam *Taiwan Database Global – Tabel TW.KA027 Statistik Kartu Kredit*. Sedangkan data pengguna kartu kredit pada tahun 2019 pada bulan Juni berjumlah 45.320.000 unit dan terdapat tiga puluh tiga lembaga penerbit kartu kredit yang beroperasi di Taiwan. (Ceicdata.com, 2018)

Peningkatan tersebut memberikan resiko terhadap krisis pembayaran penggunaan kartu kredit, yaitu adanya proporsi yang terlalu besar terhadap pemberian kartu kredit pada nasabah yang tidak memenuhi syarat. Dapat diartikan kemampuan pembayaran tagihan tidak sebanding dengan konsumsi, sehingga menyebabkan pembengkakan pada tagihan kartu kredit. Hal ini tidak hanya berdampak pada nasabah yang

akan berpengaruh terhadap kepercayaan perbankan dalam pemberian kartu kredit, melainkan lembaga keuangan atau perbankan sendiri akan mengalami hambatan dalam menjalankan kinerja bisnis perusahaan.

Manajemen resiko pada sistem keuangan atau dalam dunia perbankan memiliki peran dalam kualitas pemberian kredit dan kinerja operasional perbankan tersebut. Kualitas pemberian kredit kepada nasabah diharapkan mampu dilakukan pengembangan dan proses pemantauan yang baik agar sejalan dengan peningkatan ekonomi pasar. Manajemen resiko dilakukan untuk mengelola penilaian kredit agar pemberian kredit dapat dilakukan dengan tepat. Selain itu, pada kinerja operasional perbankan dibutuhkan adanya manajemen resiko untuk penilaian resiko yang dapat menghambat seluruh kegiatan operasional. (Yung, 2006).

Analisis terhadap prediksi resiko dilakukan dengan tujuan untuk memberikan informasi keuangan, sehingga dapat dijadikan sebagai acuan mengukur kinerja bisnis dan untuk mengurangi hambatan serta kerugian bagi lembaga keuangan. Resiko terbesar dalam dunia perbankan yaitu dalam hal pemberian kredit untuk nasabah yang memiliki potensi tidak dikembalikan pada waktu yang sesuai. Pemberian kredit yang tidak lancar diartikan sebagai kualitas kredit tergolong pada tingkat kolektabilitas tidak lancar, diragukan, dan macet. (Jamaludin, 2018).

Resiko pada pemberian kredit termasuk dalam resiko aset yang dipegang oleh nasabah sebagai pihak luar perbankan. Hal itu menyebabkan ketidakpastian dalam segi waktu pengembalian kredit yang tidak sesuai dengan perjanjian. Sedangkan, pihak perbankan menginginkan adanya perputaran aset yang produktif dan *collectable* (Jamaludin, 2018). Dengan adanya resiko tersebut, diperlukan adanya pengelolaan dan penanggulangan untuk mencegah terjadinya kredit bermasalah. Pada proses pengelolaan tersebut, adanya identifikasi terhadap karakter nasabah yang akan diberikan kepercayaan terhadap penggunaan kartu kredit perlu dilakukan. Kebutuhan informasi untuk menganalisis prediksi resiko pemberian kredit untuk meminimalisasi kesalahan penyaluran kredit kepada nasabah yang tidak memenuhi persyaratan sebagai debitur.

Upaya – upaya pencegahan kesalahan tersebut dapat dilakukan dengan mitigasi resiko penyaluran kredit. Identifikasi karakteristik nasabah dapat dilakukan sebagai salah satu kegiatan dalam mitigasi resiko penyaluran kredit. Sehingga dilakukan identifikasi karakteristik nasabah yang berhak dan memenuhi kemampuan secara finansial dalam penggunaan kartu kredit.

Beberapa penelitian yang dilakukan mengenai analisis klasifikasi kartu kredit, menggunakan algoritma *naive bayes*. Dalam penelitian tersebut diperoleh hasil analisis terhadap 23 variabel, terdapat 3 variabel yang mempengaruhi penggunaan kredit lancar dan kredit macet nasabah (Wijayanti & Sulastri, 2018). Sedangkan penelitian lainnya menggunakan algoritma metode *classification* dan *Regression Trees (CART)* untuk memodelkan *credit scoring* dalam menentukan calon nasabah pemegang kartu kredit dan pengelompokkan jenis nasabah yang sesuai dengan persyaratan pengguna kartu kredit dan nasabah yang tidak sesuai [3].

Penelitian selanjutnya [7] mengenai manajemen resiko telah dilakukan untuk mengelola aktivitas dalam dunia perbankan. Hal ini sejalan dengan kebutuhan manajemen resiko terhadap keberlangsungan segala aktivitas dalam lembaga keuangan atau perbankan. Mitigasi resiko kredit pada perbankan

telah dilakukan penelitian untuk mengetahui resiko hukum pada pemberian kredit dan mengidentifikasi kegiatan mitigasi resiko kredit oleh perbankan [5].

Berdasarkan penelitian sebelumnya yang telah dilakukan dan kebutuhan dalam identifikasi karakteristik nasabah, dapat dilakukan teknik *data mining*. Pengertian *data mining* merupakan proses penggunaan teknik statistika, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi yang dibutuhkan. (Turban, dkk., 2005 dalam [6]). Terdapat beberapa teknik yang dapat dilakukan dalam *data mining*, salah satunya adalah klasifikasi. Klasifikasi merupakan proses menemukan model atau fungsi yang menjelaskan konsep dan karakteristik suatu kelas data. [6]. Pada klasifikasi terdapat beberapa algoritma, yaitu *K-nearest neighbor*, *Logistics regression*, *Random Forest*, dan *Support Vector Machine (SVM)*.

Berdasarkan uraian diatas, permasalahan yang diangkat pada penelitian ini adalah menganalisis pengaruh faktor demografis terhadap kemungkinan *default payment* nasabah, serta menganalisis performansi beberapa metode klasifikasi untuk kemudian dilakukan penentuan metode dengan performansi terbaik untuk melakukan proses klasifikasi. Penelitian ini bertujuan untuk menentukan metode klasifikasi terbaik untuk melakukan pengelompokan nasabah berdasarkan karakteristik nasabah. Dengan penelitian ini, diharapkan dapat memberikan manfaat bahwa bank dapat lebih selektif dalam mengklasifikasikan nasabah sehingga meminimalisasi adanya kredit macet dan kerugian. Selain itu, dengan mengidentifikasi dan menganalisis faktor demografis yang memiliki pengaruh terhadap resiko *default payment*, sehingga faktor demografis dapat dipertimbangkan oleh bank dalam seleksi nasabah yang mengajukan pembuatan kartu kredit.

Penelitian ini dilakukan dengan menggunakan data nasabah sebuah bank di Taiwan selama enam bulan yaitu dari April 2005 hingga September 2005. Asumsi yang digunakan dalam melakukan pengolahan data adalah *error* data saat pengkategorian kategori usia digantikan dengan menggunakan nilai *mean* dari usia nasabah. Selain itu, *error* data saat pengkategorian kategori resiko dianggap sebagai kategori tinggi, hal ini dikarenakan bank ingIN meminimalisasi jumlah kredit bermasalah dan kemungkinan adanya *default payment* nasabah. Penelitian ini terbatas menggunakan empat metode klasifikasi yaitu *Logistic Regression*, *K Nearest neighbor*, *random forest*, dan *support vector machine*.

## Studi Literatur

Penelitian ini menggunakan beberapa studi literatur dari penelitian sebelumnya maupun *website*. Beberapa studi literatur yang digunakan adalah sebagai berikut:

### 1. Teknik *Data Mining*

*Data mining* adalah proses untuk menemukan pola dari kumpulan data set yang besar dan melibatkan teknik pembelajaran computer (*machine learning*) untuk menganalisis suatu pengetahuan atau *knowledge* secara otomatis. Adapun proses atau tahapan dari *data mining* adalah sebagai berikut :

#### 1. *Data Integration*

Merupakan tahap awal dari proses *data mining*, yaitu proses penggabungan data dari berbagai database ke dalam satu

database yang baru. Integrasi data dilakukan pada atribut – atribut yang mengidentifikasi suatu entitas yang unik seperti nama, jenis kelamin, jenis pembayaran, nomor pelanggan, dna lain sebagainya.

#### 2. *Data Selection*

Proses mengevaluasi data agar data yang digunakan sesuai dengan masalah yang akan dianalisis, karena sering kali data pada database tidak semuanya digunakan dalam proses analisisnya.

#### 3. *Data Cleaning*

Proses *cleaning* merupakan proses yang dilakukan untuk memeriksa data sehingga data yang telah diambil terhindar dari duplikasi, kesalahan cetak, atau data yang tidak konsisten.

#### 4. *Data Transformation*

Proses yang menyesuaikan data dengan format yang sesuai dengan *data mining*. Beberapa *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan.

#### 5. *Data Mining*

*Data mining* adalah proses yang paling utama pada saat metode terpilih diterapkan yaitu proses pengolahan data sehingga hasil dari proses ini adalah mendapatkan informasi dan pengetahuan yang berharga dari data yang belum diketahui sebelumnya.

#### 6. *Pattern Evaluation*

Tahap ini dilakukan untuk mengidentifikasi dan mengevaluasi pola – pola yang menarik dari informasi yang telah didapatkan setelah melalui proses *data mining*, sehingga dapat mengevaluasi hipotesa yang telah ditentukan.

#### 7. *Knowledge Presentation*

Penyajian dan visualisasi informasi dari data dan metode yang digunakan. Visualisasi dapat dilakukan untuk membantu dalam mengkomunikasikan hasil *data mining* agar mudah dimengerti oleh pembacanya.

## 2. Klasifikasi

Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. Dalam membedakan antara objek yang satu dengan yang lainnya berdasarkan atribut atau fitur data yang ada. Kelebihan Klasifikasi adalah mudah diimplementasikan dan kelemahan Klasifikasi adalah membutuhkan ketelitian yang tinggi karena keputusan akhir yang sangat dipengaruhi oleh proses klasifikasi.

### a. *Logistics Regression*

*Logistics Regression* merupakan kasus khusus model regresi linier yang menghubungkan antara satu atau beberapa variabel independen (variabel bebas) dengan variabel dependen yang berupa kategori; biasanya 0 dan 1. Tujuan dari *Logistics Regression* adalah melihat karakteristik kelompok data dan menghitung peluang responden diluar responden yang termasuk dalam penelitian, contohnya proses pengajuan kredit. Pihak bank biasanya melakukan evaluasi kelayakan seseorang layak atau tidak untuk menerima kredit pinjaman dari bank. Kelebihan dari *Logistics Regression* adalah lebih mudah digunakan secara matematis Kelemahan dari *Logistic*

*Regression* adalah rentan terhadap underfitting pada dataset yang kelasnya tidak seimbang.

b. *K-nearest neighbor* (KNN)

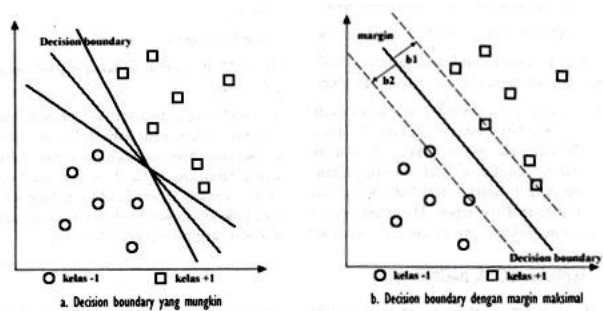
*K-nearest neighbor* merupakan teknik prediksi yang hampir sama dengan *clustering* atau pengelompokan yaitu teknik yang digunakan untuk memperkirakan apakah nilai prediksi ada dalam satu *record*, metode klasifikasi data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Termasuk dalam *supervised learning*, dimana hasil diklasifikasi berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. Kelebihan KNN adalah mudah dipahami dan diimplementasikan serta dapat menghasilkan data yang akurat sedangkan kekurangan KNN diantaranya biaya komputasi yang cukup tinggi, perlu mengetahui parameter jumlah tetangga terdekat, dan sensitif terhadap *outlier*.

c. *Random Forest*

*Random Forest* merupakan salah satu metode dalam *Decision Tree* atau pohon pengambil keputusan. *Random Forest* adalah metode yang menggunakan diagram alir yang berbentuk seperti pohon dan memiliki sebuah *root node* yang digunakan untuk mengumpulkan data, *root node* kemudian bercabang menjadi *inner node* yang berisi tentang pertanyaan tentang data dan sebuah *leaf node* yang digunakan untuk memecahkan masalah serta membuat keputusan. Kelebihan dari *Random Forest* yaitu daerah pengambilan keputusan lebih simpel dan spesifik, fleksibel dalam memilih fitur dari *inner node* yang berbeda sehingga dapat meningkatkan kualitas keputusan. Sedangkan kekurangan metode *Random Forest* yaitu kesulitan dalam mendapatkan hasil yang optimal karena hasil sangat dipengaruhi oleh desain pohon yang dibuat.

d. *Support Vector Machine*

*Support Vector Machine* (SVM) merupakan salah satu metode yang dapat digunakan dalam permasalahan klasifikasi. Konsep dasar SVM bermula dari masalah klasifikasi dua kelas sehingga membutuhkan *training set* positif dan negatif. SVM berusaha menemukan *hyperplane* (pemisah) terbaik untuk memisahkan ke dalam dua kelas dan memaksimalkan margin antara dua kelas tersebut. Pada beberapa kasus, data tidak bisa diklasifikasi menggunakan metode linier SVM, sehingga dikembangkan fungsi kernel untuk mengklasifikasikan data dalam bentuk nonlinier. SVM mempunyai akurasi klasifikasi yang lebih stabil dibandingkan dengan algoritma yang lain (Defu, Huang, Qingshan, dan Yi Jiag, 2007). Hal tersebut menjadikan SVM sebagai algoritma yang cocok untuk dijadikan solusi dalam menangani resiko kredit.

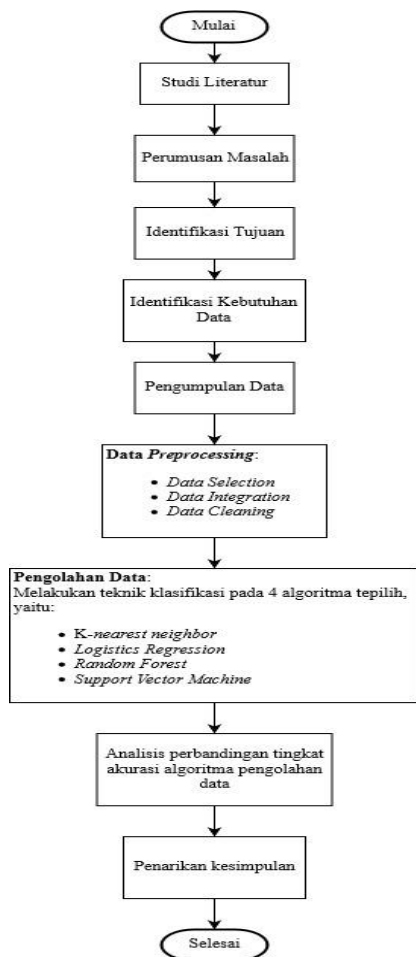


Gambar 2 Decision boundary yang mungkin untuk set data; Sumber : pemrogramanmatlab.com

Pada Gambar 2(a) terdapat beberapa pilihan *hyperplane* yang mungkin untuk set data, dan 2(b) merupakan *hyperplane* dengan margin yang paling maksimal. Meskipun 2(a) sebenarnya juga bisa menggunakan *hyperplane* sembarang, *hyperplane* dengan margin yang maksimal akan memberi generalisasi yang lebih baik pada metode klasifikasi. *Hyperplane* (batas keputusan) pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin *hyperplane* tersebut dan mencari titik maksimalnya. Garis solid pada Gambar 2(b) sebelah kanan menunjukkan *hyperplane* terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan data lingkaran dan bujur sangkar yang dilewati garis batas margin (garis putus-putus) adalah *support vector* (data yang paling dekat).

## METODE

Metode penelitian yang dilakukan berdasarkan langkah – langkah yang digambarkan pada *flowchart* di Gambar 2



**Gambar 2** Flowchart langkah – langkah penelitian

Metode penelitian yang dilakukan adalah sebagai berikut:

1. Melakukan studi literatur  
Studi literatur diperoleh dari jurnal internasional untuk mengumpulkan data – data yang dibutuhkan dalam penelitian. Studi literatur yang dilakukan terkait dengan jurnal penelitian mengenai permasalahan penyaluran kredit kepada nasabah atau calon pengguna kartu kredit.
2. Merumuskan permasalahan  
Perumusan masalah pada penelitian yaitu mengidentifikasi algoritma klasifikasi yang sesuai untuk menemukan model karakteristik nasabah yang memiliki tanggung jawab *financial* sebagai pengguna kartu kredit.
3. Mengidentifikasi tujuan  
Berdasarkan hasil studi literatur dan perumusan masalah, dilakukan identifikasi tujuan penelitian, yaitu memperoleh metode yang dapat digunakan dalam penentuan model karakteristik nasabah pengguna kartu kredit sehingga dapat meminimalisir kesalahan penyaluran kredit.
4. Mengidentifikasi kebutuhan data dan pengumpulan data  
Kebutuhan data pada penelitian ini diperoleh dari penelitian (Yeh & Lien, 2009) menggunakan obyek Negara Taiwan yang memiliki pola pengguna kartu kredit aktif yang meningkat secara signifikan. Pengumpulan data transaksi sejumlah 25.000 data pada bulan Oktober 2005 di bank penerbit uang tunai dan

kartu kredit Taiwan. Penelitian menggunakan variabel biner dengan yaitu 1: pengguna kartu kredit yang tidak kredibel dan 0: pengguna kartu kredit yang kredibel, sebagai *variable respons*. Kemudian menggunakan 23 atribut sebagai *explanatory variables*, yaitu:

X1: Jumlah kredit yang diberikan kepada nasabah  
X2: Jenis kelamin, dengan label 1: pria, 2: wanita  
X3: Tingkat Pendidikan, dengan label 1: Sekolah pascasarjana, 2: Universitas, 3: Sekolah tinggi, dan 4: lainnya

X4: Status pernikahan, dengan label 1: sudah menikah, 2: belum menikah, 3: lainnya

X5: Usia

X6 – X11: Riwayat pembayaran, dengan rincian sebagai berikut:

X6: Status pembayaran pada bulan September 2005

X7: Status pembayaran pada bulan Agustus 2005

X8: Status pembayaran pada bulan Juli 2005

X9: Status pembayaran pada bulan Juni 2005

X10: Status pembayaran pada bulan Mei 2005

X11: Status pembayaran pada bulan April 2005 dengan skala status pembayaran -2: pembayaran dilakukan lebih cepat dua bulan, -1: pembayaran tepat waktu, 0: pembayaran jaminan, 1: keterlambatan pembayaran selama satu bulan, 2: keterlambatan pembayaran selama dua bulan, 3: keterlambatan pembayaran selama tiga bulan, 4: keterlambatan pembayaran selama empat bulan, 5: keterlambatan pembayaran selama lima bulan, 6: keterlambatan pembayaran selama enam bulan, 7: keterlambatan pembayaran selama tujuh bulan, 8: keterlambatan pembayaran selama delapan bulan, 9: keterlambatan pembayaran selama sembilan bulan keatas.

X12 – X17: Jumlah tagihan, dengan rincian sebagai berikut:

X12: Jumlah tagihan pada bulan September 2005

X13: Jumlah tagihan pada bulan Agustus 2005

X14: Jumlah tagihan pada bulan Juli 2005

X15: Jumlah tagihan pada bulan Juni 2005

X16: Jumlah tagihan pada bulan Mei 2005

X17: Jumlah tagihan pada bulan April 2005

X18 – X23: Jumlah pembayaran sebelumnya, dengan rincian sebagai berikut:

X18: Jumlah yang telah dibayarkan pada bulan September 2005

X19: Jumlah yang telah dibayarkan pada bulan Agustus 2005

X20: Jumlah yang telah dibayarkan pada bulan Juli 2005

X21: Jumlah yang telah dibayarkan pada bulan Juni 2005

X22: Jumlah yang telah dibayarkan pada bulan Mei 2005

X23: Jumlah yang telah dibayarkan pada bulan April 2005

Perolehan data dibagi menjadi dua jenis kelompok, yaitu sebagai *testing data* dan *training data*. Berdasarkan perolehan data tersebut, dilakukan *data preprocessing*.

##### 5. Melakukan *data preprocessing*

Pada *data preprocessing* dilakukan untuk mengurangi kesalahan dan pencilaan terhadap data yang diperoleh sebelum dilakukan proses pengolahan data. Hal ini dilakukan dengan memilih data – data yang akan digunakan, kemudian menggabungkan data – data yang sudah dipilih dan mengidentifikasi korelasi dari data tersebut. Setelah itu, menghilangkan keberadaan *error* dan atau pencilaan jika terdapat pada data yang akan diolah. Pada *data preprocessing* menggunakan bantuan perangkat lunak *python*.

##### 6. Mengolah data

Pengolahan data dilakukan dengan teknik klasifikasi menggunakan 4 jenis algoritma, yaitu *K-nearest neighbor*, *Logistics Regression*, *Random Forest*, dan *Support Vector Machine*. Proses pengolahan data menggunakan perangkat lunak *python*.

##### 7. Analisis dan penarikan kesimpulan

Analisis dilakukan berdasarkan perbedaan tingkat akurasi pengolahan data pada masing – masing algoritma. Penyajian analisis dilakukan dengan menggunakan *visualisasi* hasil perbandingan yang mudah dipahami. Kemudian dilakukan analisis terhadap hasil pengolahan data yang memiliki tingkat akurasi yang tinggi untuk mengetahui model karakteristik nasabah pengguna kartu kredit. Berdasarkan hasil analisis tersebut dilakukan pengambilan keputusan untuk menjawab permasalahan dan tujuan penelitian.

## PENGOLAHAN DATA DAN ANALISIS HASIL

Data *default* kartu kredit nasabah di Taiwan yang diambil selama 6 bulan yaitu pada bulan April hingga bulan September 2005 dengan jumlah data sebanyak 30000. Beberapa pengolahan data dan hasil pengolahannya akan dijabarkan sebagai berikut:

### 1. *Data Preprocessing*

Pada dataset yang digunakan terdapat 25 atribut yang semuanya berupa atribut numerik sehingga tidak diperlukan adanya proses pengkategorian data. Proses *preprocessing* data dilakukan dengan tujuan untuk menghilangkan *error* dalam data yang dapat mempengaruhi keakuratan hasil penelitian. Proses ini diawali dengan membaca data dan memeriksa *error* dalam data. Berdasarkan hasil pemeriksaan *error* dan plot grafik *heatmap*, tidak didapatkan kesalahan maupun *error* dalam data awal, sehingga tidak diperlukan adanya proses *data cleaning*.

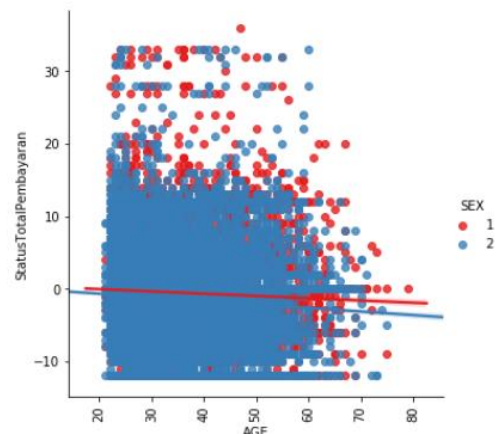
Dataset ini memiliki target kelas biner yaitu menyatakan apakah nasabah memiliki kemungkinan *default payment* pada bulan mendatang. Namun, jumlah data dengan kelas target *default* memiliki jumlah 4 kali lebih besar dibandingkan *nondefault* sehingga memungkinkan adanya bias data dan kecenderungan pada suatu kelas.

Penelitian ini memiliki tujuan untuk menganalisis seberapa besar kemungkinan *default payment* diakibatkan oleh faktor demografis serta mencari metode yang mampu membangun model untuk klasifikasi model paling akurat. Oleh karena itu, dilakukan pengkategorisasian variabel *Limit Balances*, *Usia*, *Pendidikan*, dan *Status Pernikahan*.

Variabel batas peminjaman dibagi menjadi tiga kategori yaitu nasabah dengan fasilitas *silver*, *gold*, *platinum* dengan batas pinjaman kredit yang berbeda. Variabel umur memiliki rentang data antara usia 21 tahun hingga 79 tahun yang kemudian dikategorikan menjadi tiga kategori usia yaitu *junior*, *middle*, dan *senior*. Selain kedua variabel tersebut, untuk mengetahui tingkat resiko *default payment* maka dilakukan pengkategorian variabel berdasarkan riwayat pembayaran nasabah. Kategori resiko tersebut dibagi menjadi kategori rendah, sedang, dan tinggi yang didasarkan pada jumlah status pembayaran untuk setiap bulannya. Nasabah dengan status total negatif berarti memiliki kecenderungan selalu membayar tepat waktu, sedangkan nasabah dengan total status bernilai positif memiliki kecenderungan sering melakukan penundaan pembayaran kredit.

### 2. *Data Visualization*

Setelah melakukan proses persiapan data, kemudian dilakukan proses *data visualization* yang bertujuan untuk mengubah data menjadi bentuk informasi yang mudah dipahami dan dianalisis. Dalam penelitian ini, proses visualisasi data diantaranya dilakukan untuk menggambarkan hubungan antar variabel dalam data serta merepresentasikan probabilitas *default payment* dipengaruhi oleh faktor demografis. Adapun hasil visualisasi data dapat dilihat sebagai berikut:

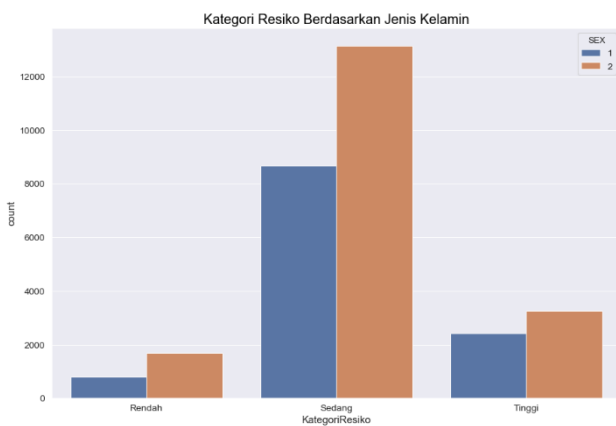


**Gambar 3** *Data Visualization* hubungan antar variabel

Gambar di atas menunjukkan hubungan antara usia dengan tingkat kategori resiko yang diwakili dengan status total pembayaran berdasarkan jenis kelamin. Berdasarkan gambar di atas, dapat dilihat bahwa pria memiliki kecenderungan lebih besar melakukan penundaan pembayaran atau *default payment*. Berdasarkan gambar dapat dilihat pula bahwa *payment default* cenderung dilakukan oleh nasabah dengan rentang usia 21 tahun hingga 55 tahun. Grafik tingkat *default payment* menurun pada rentang usia di atas 55 tahun. Hal ini membuktikan bahwa usia memiliki pengaruh yang besar terhadap probabilitas terjadinya *default payment*. Rentang usia 21 hingga 40 tahun memiliki jumlah terbanyak dalam kemungkinan *default payment* yang disebabkan berbagai faktor

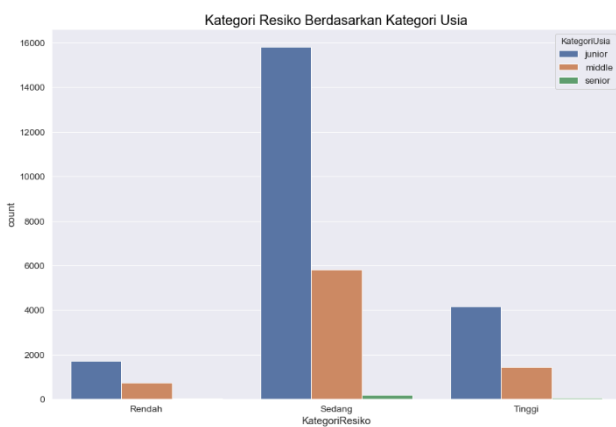
seperti tingkat kebutuhan dan pendapatan yang tidak mencukupi jumlah kredit yang dipakai.

Selain itu, visualisasi data digunakan untuk melihat pengaruh masing-masing faktor demografis terhadap tingkat kategori resiko *default payment* dari nasabah. Nasabah dengan kategori resiko rendah merupakan nasabah yang selalu membayar tagihan tepat waktu dan menunda pembayaran paling banyak selama dua bulan. Hal ini dikarenakan tujuan dari bank adalah ingin meminimasi adanya kredit macet. Toleransi sebesar 2 bulan diasumsikan sebagai batas wajar penundaan yang masih memungkinkan seorang nasabah tetap dapat dipercaya untuk melakukan kredit dan transaksi. Berikut ini merupakan grafik yang menggambarkan masing-masing faktor demografis dengan kategori resiko:



**Gambar 4** Grafik tingkat resiko berdasarkan jenis kelamin nasabah

Gambar di atas menggambarkan tingkat kategori resiko berdasarkan jenis kelamin nasabah. Berdasarkan gambar di atas, untuk masing-masing kategori resiko, jenis kelamin pria memiliki jumlah *default payment* yang lebih tinggi dibandingkan wanita.

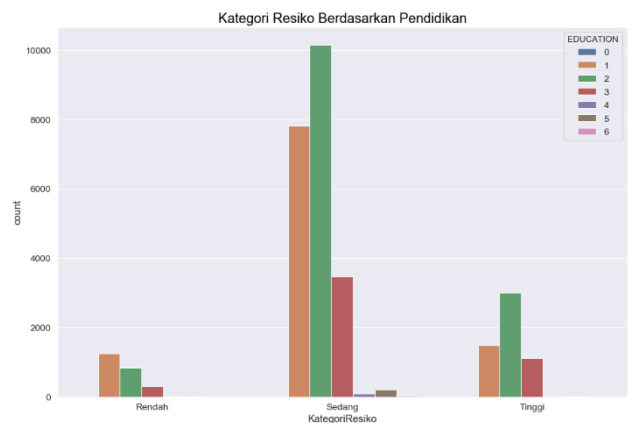


**Gambar 5** Grafik tingkat resiko berdasarkan kategori usia

Gambar di atas menggambarkan tingkat kategori resiko berdasarkan kategori usia. Kategori usia yang dimaksudkan adalah tiga kategori usia seperti

yang sudah dijelaskan pada subbab sebelumnya. Kategori usia tersebut yaitu kategori *junior* yang memiliki rentang usia antara 21-40 tahun. Kategori *middle* merupakan nasabah dengan rentang usia antara 40-60 tahun, sedangkan kategori usia *senior* terdiri dari nasabah dengan rentang usia 60-80 tahun. Berdasarkan kategori usia tersebut, dapat dilihat bahwa kecenderungan terjadinya *default payment* untuk semua jenis kategori resiko diwakili oleh kategori usia *junior*. Sedangkan kategori usia menengah menempati posisi kedua penyebab kemungkinan terjadinya *default payment* di semua kategori. Kategori senior menjadi kategori paling jarang atau paling minimal probabilitas terjadinya *default payment* untuk semua kategori resiko.

Grafik menggambarkan bahwa kategori usia *junior* memiliki jumlah paling tinggi dan probabilitas terbesar menyebabkan adanya kredit macet. Hal ini dikarenakan kategori junior berada pada rentang usia 21 hingga 40 tahun yang dapat dikatakan sebagai usia produktif. Pada usia tersebut, nasabah cenderung memanfaatkan kartu kredit untuk memenuhi kebutuhan namun tidak diimbangi dengan penghasilan yang mencukupi dan komitmen melunasi kredit yang tinggi. Oleh karena itu, faktor usia dapat dijadikan dasar acuan pertimbangan dalam menentukan kelayakan pinjaman oleh nasabah.

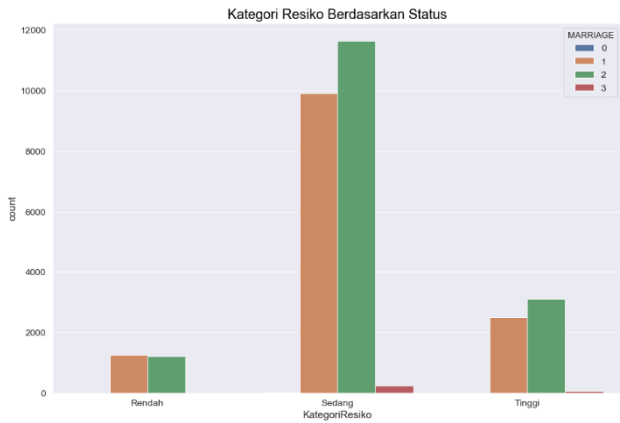


**Gambar 6** Grafik kategori resiko berdasarkan tingkat pendidikan nasabah

Gambar di atas merupakan grafik kategori resiko berdasarkan tingkat pendidikan nasabah. Berdasarkan persebaran kategori resiko berdasarkan pendidikan, tingkat resiko sedang menjadi yang paling tinggi dibandingkan tingkat resiko lain. Kategori pendidikan 1, 2, dan 3 mewakili jumlah terbanyak pada masing-masing kategori resiko. kategori pendidikan dengan indeks 1 adalah nasabah yang telah lulus perguruan tinggi, kategori indeks 2 digunakan untuk nasabah yang masih berstatus mahasiswa, sedangkan indeks 3 mewakili nasabah yang masih duduk di bangku menengah atas.

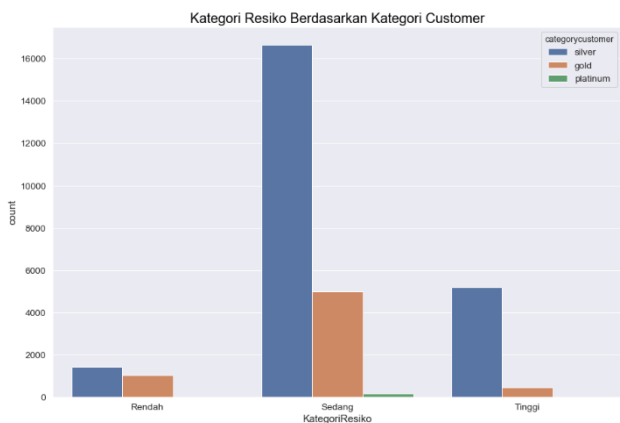
Pada kategori resiko rendah, nasabah yang sudah lulus menjadi kelompok tertinggi yang menyebabkan adanya *default payment*. Kategori

selanjutnya yaitu kategori mahasiswa dan terendah adalah kategori siswa sekolah menengah. Selanjutnya, pada kategori resiko sedang menunjukkan jumlah yang sangat tinggi dibandingkan dengan kategori resiko rendah dan tinggi dimana kategori mahasiswa memiliki probabilitas terbesar menunda pembayaran kredit. Sedangkan kategori tinggi juga diwakili oleh kategori mahasiswa sebagai kelompok tertinggi yang memiliki kategori resiko tinggi dalam melakukan *default payment*.



**Gambar 7** Grafik kategori resiko berdasarkan status perkawinan

Gambar di atas menunjukkan grafik kategori resiko berdasarkan status perkawinan. grafik menggambarkan bahwa berdasarkan status pernikahan kecenderungan resiko *default payment* terbesar adalah kategori resiko sedang. Grafik berwarna hijau menunjukkan status perkawinan masih lajang. Berdasarkan grafik dapat disimpulkan bahwa status perkawinan lajang memiliki probabilitas tertinggi yang mempengaruhi tingka *default payment*.

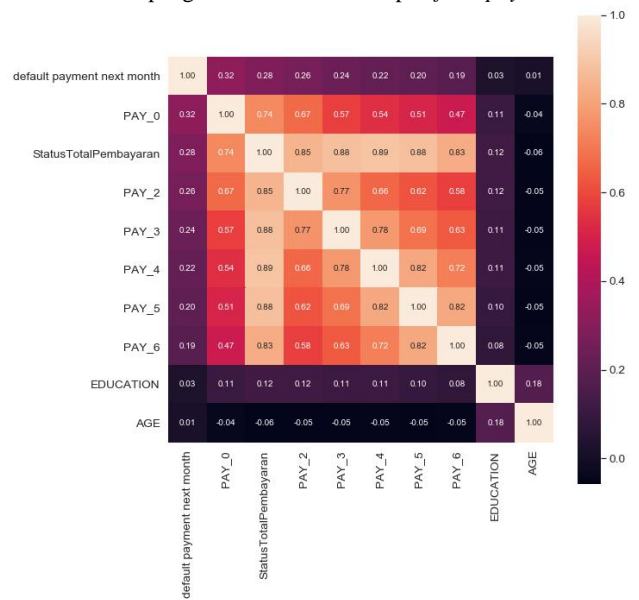


**Gambar 8** Grafik kategori resiko berdasarkan kategori customer

Gambar di atas menunjukkan grafik kategori resiko berdasarkan kategori customer. Kategori customer yang dimaksudkan adalah kategori customer berdasarkan batas maksimum peminjaman seperti yang telah dikategorikan pada tahap *preprocessing*. Kategori customer dibagi menjadi 3 kelas fasilitas yaitu kelas silver, gold, dan platinum. Kategori tersebut dibedakan

dengan fasilitas yang didapatkan oleh nasabah dimana kategori platinum memiliki keuntungan mendapatkan batas maksimal peminjaman yang terbesar dibandingkan dengan kategori lainnya. Berdasarkan grafik dapat disimpulkan bahwa untuk setiap kategori resiko, nasabah dengan fasilitas kartu silver memiliki kemungkinan terbesar melakukan penundaan pembayaran kredit yang mengakibatkan kredit macet.

Berdasarkan grafik-grafik di atas, dapat dilihat pengaruh dan perbandingan masing-masing faktor demografis terhadap *payment default*. Pada dasarnya pengaruh terbesar terhadap *payment default* tidak dipengaruhi secara langsung oleh faktor demografis. Hal ini dikarenakan dasar perhitungan dan penentuan kemungkinan *default payment* didasarkan pada riwayat pembayaran kredit nasabah. Namun, faktor demografis dapat dijadikan bahan pertimbangan bagi bank saat melakukan seleksi awal pengajua kartu kredit dan penentuan batas peminjaman maksimal nasabah. Berikut ini merupakan 10 fator utama yang memiliki pengaruh terbesar terhadap *default payment*:



**Gambar 9** Faktor utama pengaruh *default payment*

3. Penentuan Model Klasifikasi

Klasifikasi bertujuan untuk membangun suatu model yang dapat mengidentifikasi dan mengelompokan nasabah berdasarkan kriteria tertentu ke dalam kelompok yang telah ditetapkan. Klasifikasi termasuk ke dalam metode *supervised learning* sehingga proses pembangunan model dibutuhkan dengan mengolah data *training* untuk membangun sebuah model lalu dilakukan validasi pada data *testing*. Berikut ini beberapa metode klasifikasi yang digunakan untuk menentukan *default payment* nasabah bank:

- a) Perhitungan akurasi algoritma *Logistics Regression*:



```

precision    recall  f1-score   support

 0         0.78    1.00    0.88    7058
 1         0.00    0.00    0.00    1942

 accuracy          0.78    9000
 macro avg         0.39    0.50    0.44    9000
 weighted avg      0.61    0.78    0.69    9000

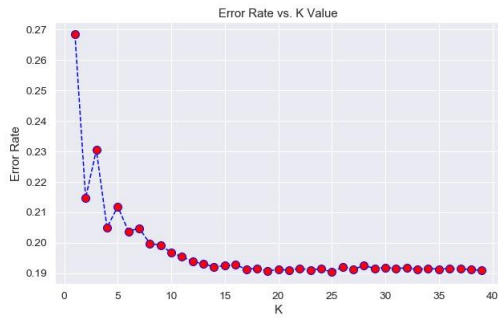
 [[7056 1942]
 [ 2    0]]

 Accuracy Score for model1: 0.784

 F1-Score for Model: 0.0
    
```

**Gambar 10** Perhitungan akurasi algoritma *Logistics Regression*

b) Perhitungan akurasi algoritma *K-nearest neighbor*



**Gambar 11** Grafik akurasi algoritma *K-nearest neighbor*

WITH K=40

```

precision    recall  f1-score   support

 0         0.79    0.98    0.88    7058
 1         0.46    0.07    0.12    1942

 accuracy          0.78    9000
 macro avg         0.63    0.52    0.50    9000
 weighted avg      0.72    0.78    0.71    9000

 Accuracy Score for model1: 0.7814444444444445
 [[6893 165]
 [1802 140]]

 F1-Score for Model: 0.12461059190031153
    
```

**Gambar 12** Perhitungan akurasi algoritma *K-nearest neighbor*

c) Perhitungan akurasi algoritma *Random Forest*

Selain menggunakan algoritma *logistics regression* dan *k-nearest neighbor*, metode lain yang digunakan adalah dengan menggunakan metode *random forest*. Pengolahan dengan metode ini sama dengan metode sebelumnya yaitu dengan membagi data menjadi data *training* dan data *testing* lalu dilakukan pembangunan model dan prediksi model. Setelah itu dilakukan perhitungan terhadap tingkat presisi, akurasi, dan *f1\_score* seperti yang dapat dilihat pada gambar di bawah ini:

```

precision    recall  f1-score   support

 0         0.83    0.95    0.89    6946
 1         0.66    0.35    0.46    2054

 accuracy          0.81    9000
 macro avg         0.75    0.65    0.67    9000
 weighted avg      0.79    0.81    0.79    9000

 [[6579 367]
 [1334 720]]

 Accuracy Score for model1: 0.811

 F1-Score for Model: 0.4584527220630373
    
```

**Gambar 13** Perhitungan akurasi algoritma *Random Forest*

Gambar di atas menunjukkan hasil prediksi dengan menggunakan metode *random forest*. Berdasarkan data di atas, dapat dilihat bahwa tingkat akurasi metode *random forest* lebih tinggi dibandingkan metode *logistics regression* dan *k nearest neighbor*. Selain itu, nilai *f1 score* pada metode ini juga memiliki nilai yang lebih tinggi dibandingkan dengan dua metode sebelumnya.

d) Perhitungan akurasi algoritma *Support Vector Machine*

Pengolahan dengan metode SVM sama dengan metode sebelumnya yaitu dengan membagi data menjadi data *training* dan data *testing* lalu dilakukan pembangunan model dan prediksi model. Setelah itu dilakukan perhitungan terhadap tingkat presisi, akurasi, dan *f1\_score* seperti yang dapat dilihat pada gambar di bawah ini:

```

precision    recall  f1-score   support

 0         1.00    1.00    1.00    7742
 1         1.00    1.00    1.00    2158

 accuracy          1.00    9900
 macro avg         1.00    1.00    1.00    9900
 weighted avg      1.00    1.00    1.00    9900

 [[7742  0]
 [ 0 2158]]

 Accuracy Score for model1: 0.9998989898989899

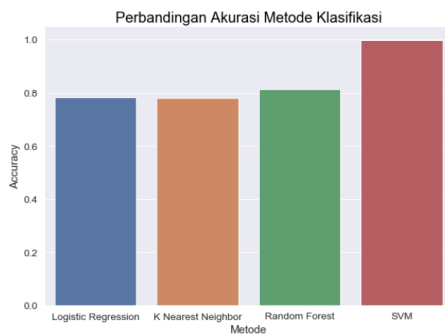
 F1-Score for Model: 0.9997682502896872
    
```

**Gambar 14** Perhitungan akurasi algoritma *Support Vector Machine*

Berdasarkan gambar di atas, dapat diketahui bahwa tingkat akurasi dan nilai *f1 score* dari metode SVM jauh lebih tinggi dibandingkan metode-metode sebelumnya.

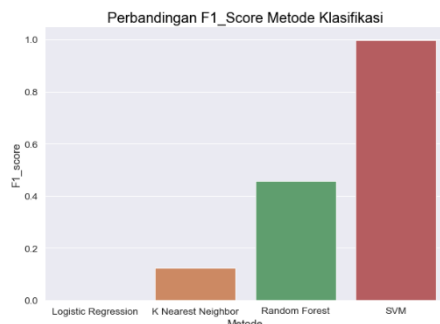
e) Perbandingan Keseluruhan Akurasi Metode Klasifikasi

Setelah melakukan percobaan dengan menggunakan empat metode klasifikasi, didapatkan hasil akurasi untuk masing-masing metode. Metode dengan tingkat akurasi paling tinggi yaitu metode SVM dengan tingkat akurasi 0,998. Grafik perbandingan tingkat akurasi masing-masing metode klasifikasi dapat dilihat seperti di bawah ini.



**Gambar 15** Perbandingan Keseluruhan Akurasi Metode Klasifikasi

Selain membandingkan performansi model dengan menggunakan tingkat akurasi, performansi masing-masing model juga dapat dibuktikan dengan menggunakan nilai *f1 score*. Nilai *f1 score* digunakan karena nilai *f1 score* merupakan rata-rata bobot dari presisi dan akurasi. Nilai ini memperhitungkan kemungkinan *default* dan *nondefault*. Nilai *f1 score* digunakan untuk menanggulangi kesalahan akurasi karena nilai akurasi lebih optimal jika digunakan untuk nilai *default* dan *non default* yang sebanding. Sehubungan dengan data nasabah yang memiliki kecenderungan pada satu kelas, nilai *f1 score* diperhitungkan untuk membuktikan performansi model yang dipilih. Berdasarkan pengolahan data, nilai *f1 score* tertinggi dihasilkan oleh metode SVM. Grafik perbandingan nilai *f1 score* dapat dilihat sebagai berikut.



**Gambar 16** Grafik perbandingan nilai *f1 score*

## KESIMPULAN

Berdasarkan pengolahan data menggunakan beberapa metode klasifikasi di atas, dapat disimpulkan bahwa metode klasifikasi paling tepat untuk mengklasifikasikan *default payment* nasabah bank adalah dengan menggunakan metode *support vector machine*. Hal ini didasarkan pada tingkat akurasi dan nilai *f1 score* dari metode SVM yang lebih tinggi dibandingkan dengan metode lainnya.

Berdasarkan hasil analisis data, didapatkan kesimpulan bahwa dasar penentuan *default payment* nasabah adalah riwayat pembayaran kredit pada bulan-bulan sebelumnya. Status pembayaran setiap bulan memiliki korelasi yang tinggi terhadap penentuan *default payment* nasabah. Oleh karena itu, untuk menentukan apakah nasabah masih kredibel untuk melakukan

peminjaman, bank perlu memperhatikan dan mempertimbangkan riwayat pembayaran nasabah. Selain itu, faktor demografis memiliki dampak tidak langsung terhadap risiko terjadinya kredit macet oleh nasabah. Oleh karena itu, bank perlu mempertimbangkan faktor demografis calon nasabah sebagai pertimbangan dalam pengajuan kartu kredit dan penentuan jumlah maksimum pinjaman.

## REFERENSI

- [1] B. Waluyo, Dedy dan J. Agung, "Menjaga Stabilitas, Mendorong Reformasi Struktural untuk Pertumbuhan Ekonomi yang Berkelanjutan," 2013.
- [2] V. Gunawan, "Studi Tnetang Pengaruh Nilai Pelanggan Inti dan Peripheral Terhadap Minat Menggunakan Kartu Kredit untuk Meningkatkan Keputusan Menggunakan Kartu Kredit," 2017.
- [3] R. Yunindya, A. Kudus dan T. S. Yanti, "Model Credit Scoring Menggunakan Metode Classification dan Regression Trees (CART) pada Data Kartu Kredit," *Prosiding Statistika*, 2016.
- [4] S. Yung, "Manajemen Resiko dalam Dunia Perbankan," *Jurnal Sistem Informasi UKM*, vol. 1, pp. 63-71, 2006.
- [5] F. Jamaludin, "Mitigasi Resiko Kredit Perbankan," *Journal of Islamic Economic Law*, 2018.
- [6] R. Wijayanti dan Sulastri, "Analisa Klasifikasi Kartu Kredit Menggunakan Algoritma Naive Bayes," *Prosiding Sintak*, 2018.
- [7] L. K. Sari, "Penerapan Manajemen Resiko pada Perbankan Indonesia".
- [8] I.-C. Yeh dan C.-h. Lien, "The Comparisons of Data Mining Techniques for The Predictive Accuracy of Probability of Default of Credit Card Clients," *Expert System with Application*, pp. 2473-2480, 2009.
- [1] Ceicdata.com. (2018). Taiwan Credit Card: Cards Activated. Diakses pada 28 November 2019. Dari <https://www.ceicdata.com/en/taiwan/credit-card-statistics/credit-card-cards-activated>
- [2] Banking Bureau. (2019, 01 Agustus). Statistics of Credit Card & Cash Card & Electronic Stored Value Cards Business Operation as of January 2018. Diakses pada 28 November 2019. Dari [https://www.banking.gov.tw/en/home.jsp?id=30&parentpath=0,3&mcustomize=multimessage\\_view.jsp&dataserno=201908010001&aplistdn=ou=news,ou=multisite,ou=english,ou=ap\\_root,o=fsc,c=tw&dttable=News](https://www.banking.gov.tw/en/home.jsp?id=30&parentpath=0,3&mcustomize=multimessage_view.jsp&dataserno=201908010001&aplistdn=ou=news,ou=multisite,ou=english,ou=ap_root,o=fsc,c=tw&dttable=News)
- [3] Lykra. (2018, 23 Juli). Mengenal Decision Tree dan Manfaatnya. Diakses pada 28 November 2019. Dari <https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d>
- [4] Kadir, Wa Nurhayah, dkk. (2019, Juni). Penerapan Data Mining Dengan Metode Knearest Neighbor (KNN) Untuk Mengelompokan Minat Konsumen Asuransi (PT. Jasaraharja Putera). Diakses pada 28 novembraer 2019. Dari <http://ojs.uho.ac.id/index.php/semantik>
- [5] Prasetyowati, Yosi Ghea. (6 Juni). Klasifikasi menggunakan Metode KNN (K-Nearest Neighbor) dalam Python. Diakses pada 28 November 2019. Dari

<https://medium.com/@16611130/klasifikasi-menggunakan-metode-knn-k-nearest-neighbor-dalam-python-a40e79a74101>

- [6] Pamungkas, Adi. 2016. Support Vector Machine. Diakses pada 28 November 2019. Dari <https://pemrogramanmatlab.com/data-mining-menggunakan-matlab/support-vector-machine-svm-menggunakan-matlab/>
- [7] Wihandika, Randy, Dian Eka Ratnawati. 2018. Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. Universitas Brawijaya, Malang.
- [8] Han, J. Kamber, M & Jian, Pei. 2011. Dalam jurnal Nasional Teknologi dan Sistem Informasi. Penerapan Data Mining dalam Pengelompokan penderita Thalassaemia. Universitas Siliwangi, Tasikmalaya.
- [9] Wu, X dan Kumar, V. 2009. Dalam Jurnal Nasional Teknologi dan Sistem Informasi. Penerapan Data Mining dalam Pengelompokan penderita Thalassaemia. Universitas Siliwangi, Tasikmalaya.
- [10] Defu Zhang, Huang Hongyi, Qingshan Chen, and Yi Jiang. 2007. Dalam Jurnal Teknologi Informasi. Optimasi Algoritma Support Vector Machine (SVM) Menggunakan Adaboost Untuk Penilaian Risiko Kredit. Universitas Dian Nuswantoro, Semarang.
- [11] Han, J, Micheline K, Jian Pei. 2012. Dalam jurnal Implementasi Algoritma Genetika Untuk Optimalisasi Random Forest Dalam Proses Klasifikasi Penerimaan Tenaga Kerja Baru Studi Kasus PT. XYZ. Institut Teknologi Sepuluh Nopember, Surabaya
- [12] Chien, C.F. dan L.F. Chen. 2008. Dalam jurnal Implementasi Algoritma Genetika Untuk Optimalisasi Random Forest Dalam Proses Klasifikasi Penerimaan Tenaga Kerja Baru Studi Kasus PT. XYZ. Institut Teknologi Sepuluh Nopember, Surabaya