Penggunaan Algoritma Forest, SVM, dan Decision Tree dalam Memprediksi Peminjaman Kredit Pada Bank

Jacintha Cordelie
Universitas Multimedia Nusantara
Banten, Indonesia
jacintha.cordelie@student.umn.ac.id

***Abstrak*— Dalam penelitian ini, peneliti akan membuat model untuk melakukan prediksi segmentasi peminjaman kredit pada suatu bank dengan menggunakan tiga jenis algoritma, yaitu Forest, *Support Vector Machine* atau SVM, dan Decision Tree, dengan menggunakan SAS Cloud Analytical Services. Hasil model dari ketiga algoritma ini akan dipaparkan pada artikel ilmiah ini, dan akan dijelaskan dengan visualisasi. Kerangka kerja yang digunakan oleh peneliti adalah *Cross-Industry Standard Process for Data Mining* atau yang biasanya dikenal sebagai CRISP-DM. Data yang digunakan dalam penelitian ini adalah data peminjaman kredit pada suatu bank. Di dalam data ini, terdapat 1000 baris dan 10 kolom. Dan dari hasilnya, didapat bahwa Model Decision Tree merupakan model yang terbaik untuk kasus prediksi ini karena: 1) Model ini memiliki *misclassification rate* yang paling kecil dari antara kedua model lainnya, 2) Model ini memiliki nilai cumulative lift yang cukup baik untuk hampir sebagian besar persentase data, 3) Model ini memiliki nilai *sensitivity* dan 1-*specificity* yang terbaik dari antara kedua model lainnya.**

***Kata Kunci*— CRISP-DM; SVM; Random Forest; Decision Tree**

Abstract*— *In this research, the researcher will develop a model to predict credit loan segmentation in a bank using three types of algorithms: Random Forest, Support Vector Machine (SVM), and Decision Tree. The SAS Cloud Analytical Services will be utilized for this purpose. The results of the models generated by these three algorithms will be presented in this scientific article and explained through visualizations. The researcher will adopt the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The dataset used in this study consists of 1000 rows and 10 columns, containing credit loan data from a bank. Based on the findings, it is concluded that the Decision Tree model is the best model for this prediction task due to the following reasons: 1) This model exhibits the lowest misclassification rate compared to the other two models, 2) The Decision Tree model demonstrates a satisfactory cumulative lift value for the majority of the data percentages, 3) The Decision Tree model achieves the best values for sensitivity and 1-specificity among the other two models.

Keywords*— *CRISP-DM; SVM; Random Forest; Decision Tree

I. LATAR BELAKANG DAN PEMBAHASAN BISNIS

Dalam beberapa dekade terakhir, banyak yang melakukan peminjaman yang berujung tidak dikembalikan. Sudah banyak sekali peminjam yang kabur setelah meminjam, dan tidak bertanggung jawab untuk mengembalikan pinjaman yang sudah dipinjamnya. Hal ini menjadikan bisnis kredit pinjaman sebagai bisnis yang sangat rawan akan kerugian.

Namun, kemajuan teknologi informasi dan komunikasi telah memberikan dampak yang signifikan bagi sektor keuangan, termasuk industri pinjaman kredit. Dalam era digital yang terus berkembang ini, lembaga keuangan dan perusahaan *fintech* telah bergeser menuju metode yang lebih canggih dalam mengevaluasi risiko kredit dan membuat keputusan yang lebih cerdas dalam memberikan pinjaman kepada individu dan perusahaan. Salah satu pendekatan yang menjanjikan dalam meningkatkan akurasi dan efisiensi proses penilaian kredit adalah penerapan teknik *machine learning*.

Penelitian mengenai prediksi pinjaman kredit dengan menggunakan *machine learning* memiliki potensi besar untuk mengoptimalkan proses penilaian kredit yang ada saat ini. Dengan menggunakan *machine learning*, bank dapat menemukan tujuan yang sebenarnya dari peminjam, dan meningkatkan pengambilan keputusan apakah peminjaman perlu disetujui atau tidak.

Selain itu, penggunaan *machine learning* juga dapat membantu mengatasi beberapa tantangan yang dihadapi dalam penilaian kredit secara manual. Sebagai contoh, *machine learning* dapat mengurangi subjektivitas dan bias yang mungkin muncul dalam proses penilaian manual. Dengan menggunakan algoritma yang objektif dan berdasarkan data, *machine learning* dapat menghasilkan penilaian yang lebih adil dan netral.

II. TINJAUAN LITERATUR

Pada penelitian ini, terdapat beberapa kajian literatur yang berisikan literasi terkait metode penelitian yang digunakan, beserta algoritma-algoritma yang akan digunakan, seperti CRISP-DM, Random Forest atau Forest, SVM, dan Decision Tree.

A. CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah framework yang banyak digunakan untuk proyek data mining dan machine learning. CRISP-DM memberikan tahapan terstruktur untuk membantu proses pengembangan, pengujian, dan penerapan model prediksi.

Berikut langkah-langkah dalam menjalankan proses big data analytics di bidang perbankan dengan menggunakan CRISP-DM adalah sebagai berikut:

1. *Business Understanding*

Tahap pertama pada *framework* CRISP-DM adalah memahami permasalahan masalah atau tujuan yang ingin dicapai oleh proyek. Ini melibatkan identifikasi tujuan proyek dan menentukan indikator kinerja utama (KPI) yang akan digunakan untuk mengukur keberhasilan.

2. *Data Understanding*

Tahap selanjutnya adalah mengumpulkan dan menganalisis data yang akan digunakan. Langkah ini termasuk identifikasi sumber data, memeriksa kualitas data, dan melakukan analisis data eksplorasi untuk mendapatkan wawasan tentang data.

3. *Data Preparation*

Tahap berikutnya adalah menyiapkan data untuk dianalisis dengan membersihkan, mengubah, dan memanipulasinya seperlunya. Ini melibatkan penghapusan data yang hilang atau duplikat, pengkodean variabel kategori, dan normalisasi variabel numerik.

4. *Modeling*

Setelah data siap untuk dibuat model, tim dapat memulai untuk membuat model. Pada fase ini, terdapat beberapa hal yang harus dilakukan, yaitu: memilih teknik pemodelan, membangun model, melakukan perbandingan model.

5. *Evaluation*

Pada tahap ini, data analyst melakukan evaluasi kinerja model yang telah dibuat pada data baru. Pada tahap *evaluation* terdapat pengujian model menggunakan menggunakan teknik *cross-validation* untuk memastikan bahwa model tidak *overfitting*.

6. *Deployment*

Tahap terakhir dari CRISP-DM adalah mengintegrasikannya ke dalam proses bisnis. Langkah ini melibatkan pembuatan API atau *interface* lain agar model dapat digunakan oleh aplikasi, dan dapat dilakukan pemantauan kinerja dari waktu ke waktu untuk memastikan bahwa model yang dibuat selalu memberikan prediksi yang akurat.

B. Random Forest (Forest)

Breiman dalam karya ilmiah oleh Religa mendefinisikan Algoritma Random Forest sebagai algoritma yang menggunakan metode pemisahan biner secara berulang untuk mencapai simpul-simpul akhir dalam struktur pohon berdasarkan pada pohon klasifikasi dan regresi. Yoo et. al dalam karya ilmiah yang sama juga menyatakan bahwa Algoritma Random Forest melakukan klasifikasi dengan menggabungkan beberapa pohon dan menggunakan mayoritas kemunculan untuk membuat keputusan akhir.. Data *training* pada algoritma ini diformulasikan sebagai berikut:

$$S=\{(x_i,\ y_j),\ i=1,\ 2,\ ...,N;\ j=1,\ 2,\ ...,\ M\}$$

Dengan :

- x = sampel
- y = variabel fitur S
- N = jumlah sampel *training*
- M = variabel fitur pada setiap sampel[1].

C. SVM

Suryati et. al dalam karya ilmiahnya mendefinisikan Algoritma SVM sebagai sebuah metode pembelajaran mesin yang bertujuan untuk menemukan *hyperplane* terbaik yang dapat memisahkan dua kelas dalam ruang input. Algoritma klasifikasi SVM menggunakan data *training* untuk membentuk model klasifikasi, dan model tersebut digunakan untuk memprediksi kelas dari data baru yang belum pernah dilihat sebelumnya, yang disebut data pengujian. [2] Support Vector Machine (SVM) dapat melakukan klasifikasi pada data yang dapat dipisahkan secara linear maupun non-linear.[3] Dalam SVM Linear, setiap data *training* dikenal sebagai (xi, yi), di mana i = 1,2,...,N dan xi={xi1,xi2,...,xiq} T merupakan atribut untuk data *training* ke-i, sementara yi bernilai {-1, +1} sebagai label kelasnya. [4]

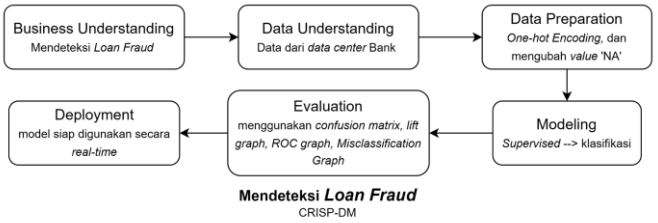
D. Decision Tree

Decision Tree adalah metode prediksi dan klasifikasi yang populer dan kuat, yang mengubah fakta menjadi pohon keputusan untuk merepresentasikan aturan. [5] Dalam proses pembuatan Decision Tree, perlu dihitung nilai *entropy*. *Entropy* merupakan ukuran *impurity* dari suatu atribut. Semakin rendah nilai *entropy*, semakin baik atribut tersebut untuk mengekstraksi kelas. Rumus untuk menghitung *entropy* adalah sebagai berikut. [6]

$$Entropy(S) = -\sum_{i=1}^{n} pi * log_2 pi$$

III. METODOLOGI PENELITIAN

Dalam melakukan penelitian ini, peneliti menggunakan kerangka kerja CRISP-DM yang terdiri atas *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.



Gambar 1 (*Flowchart Framework* CRISP-DM dalam mendeteksi *loan faoud*)

A. Business Understanding

Pada tahap yang pertama ini, peneliti menentukan tujuan dari penelitian ini, dimana peneliti ingin mendeteksi penipuan pinjaman pada suatu bank. Selain itu, peneliti juga menetapkan ekspektasi hasil keluaran dari penelitian ini, dimana peneliti ingin menemukan pola untuk membedakan pinjaman mana yang merupakan penipuan dan mana yang tidak.

B. Data Understanding

Pada tahap ini, peneliti akan memahami data dengan menggunakan metode *Exploratory Data Analysis* (EDA), dimana EDA ini merupakan merupakan pada serangkaian teknik dan pendekatan yang digunakan untuk memahami dan menganalisis data secara sistematis sebelum menjalankan proses analisis yang lebih mendalam. EDA bertujuan untuk menemukan pola-pola, hubungan, anomali, dan wawasan baru dalam sebuah *big data*. EDA membantu para peneliti untuk mengidentifikasi karakteristik penting, mengatasi kehilangan atau data yang tidak lengkap, mengidentifikasi outliers, dan menjelaskan hubungan antara variabel dalam dataset. Sehingga, EDA sangat membantu dalam merencanakan pendekatan analisis yang lebih lanjut dan memastikan bahwa data yang digunakan dapat dipercaya.

Sebelum memulai proses EDA, peneliti akan menjelaskan isi dari dataset, berserta penjelasan dari setiap kolomnya.

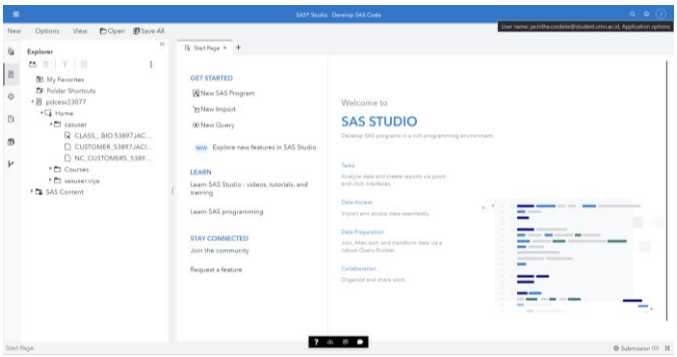
1. Kolom tanpa nama yang berisi nomor baris data (numerik). Kolom ini tidak akan digunakan pada penelitian ini,
2. Age, yang menjelaskan usia peminjam (numerik),
3. Sex, yang mendeskripsikan jenis kelamin peminjam (kategorikal), yang memiliki 2 value:
   a. male → peminjam berjenis kelamin laki-laki,
   b. female → peminjam berjenis kelamin perempuan.
4. Job, yang mendeskripsikan pekerjaan peminjam (kategorikal, pada SAS Studio akan terbaca sebagai numerik). Kolom ini dilakukan encoding, bawaan dari sumber data dan memiliki 4 value, yaitu:
   a. 0 → unskilled and non-resident (tidak terampil dan bukan penduduk),
   b. 1 → unskilled and resident (tidak terampil dan merupakan penduduk),
   c. 2 → skilled (terampil),
   d. 3 → highly skilled (sangat terampil).
5. Housing, yang menjelaskan status tempat tinggalnya (kategorikal), yang memiliki 3 value:
   a. own → rumah tempat tinggal peminjam merupakan rumah pribadi,
   b. rent → rumah tempat tinggal peminjam merupakan rumah sewaan,
   c. free → rumah tempat tinggal peminjam merupakan akomodasi gratis.
6. Saving accounts, yang menjelaskan jumlah uang dalam tabungan (kategorikal), yang memiliki 4 value:
   a. little → uang dalam tabungan peminjam berjumlah sedikit,
   b. moderate → uang dalam tabungan peminjam berjumlah sedang,
   c. quite rich → uang dalam tabungan peminjam berjumlah lumayan banyak,
   d. rich → uang dalam tabungan peminjam berjumlah banyak,
7. Checking accounts, yang menjelaskan jumlah rekening cek peminjam (dalam satuan Deutsche Mark) (kategorikal), yang memiliki 4 value:
   a. little → rekening cek peminjam berjumlah sedikit,

b. moderate → rekening cek peminjam berjumlah sedang,
c. quite rich → rekening cek peminjam berjumlah lumayan banyak,
d. rich → rekening cek peminjam berjumlah banyak,

8. Credit amount, yang menjelaskan jumlah uang yang dipinjam (dalam satuan Deutsche Mark) (numerik),
9. Duration, yang menjelaskan lama uang tersebut dipinjam (kategorikal),
10. Purpose, yang menjelaskan tujuan peminjaman uang (kategorikal), yang memiliki 8 value:
    a. car → untuk keperluan berkaitan dengan mobil,
    b. furniture/equipment → untuk keperluan berkaitan dengan furnitur,
    c. radio/TV → untuk keperluan berkaitan dengan radio atau TV,
    d. domestic appliances → untuk keperluan berkaitan dengan peralatan rumah tangga,
    e. repairs → untuk keperluan perbaikan,
    f. education → untuk keperluan berkaitan dengan edukasi,
    g. business → untuk keperluan berkaitan dengan bisnis,
    h. vacation/others → untuk keperluan berkaitan dengan liburan.

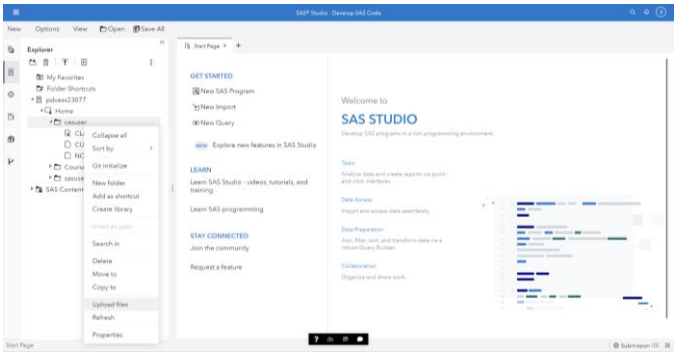
Selanjutnya, peneliti akan menjelaskan statistik dari *dataset* ini dengan menggunakan EDA. Nama sebelumnya, peneliti akan menjelaskan tahap-tahap yang dilakukan peneliti untuk meng-import data, yaitu sebagai berikut:

1. Pertama-tama peneliti masuk ke SAS Viya dengan menggunakan *e-mail* dan *password* akun yang sudah dibuat, kemudian peneliti masuk ke menu Develop SAS Code. Tampilan menu tersebut adalah sebagai berikut:



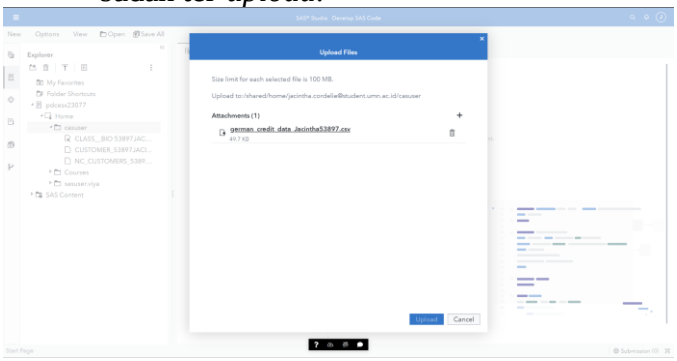
Gambar 2 (Tampilan menu Develop SAS Code)

2. Berikutnya, peneliti akan *extend folder* pdcesx23077 → Home → casuser, lalu klik kanan pada *folder* casuser, dan pilih "Upload files"



Gambar 3 (Tampilan untuk *extend folder*)

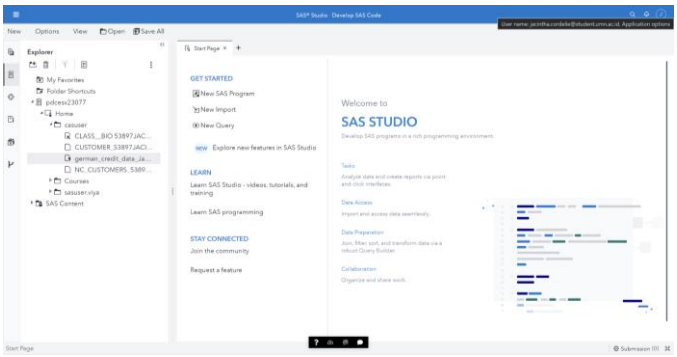
3. Selanjutnya, peneliti akan mengupload *file* yang ingin digunakan untuk membuat model dari penyimpanan perangkat peneliti. Pada kasus ini, saya akan mengupload *file* yang bernama "GERMAN_CREDIT_DATA_JACINTHA 53897.csv". Berikut tampilannya ketika *file* sudah ter-*upload*.



Gambar 4 (Tampilan untuk *upload file*)

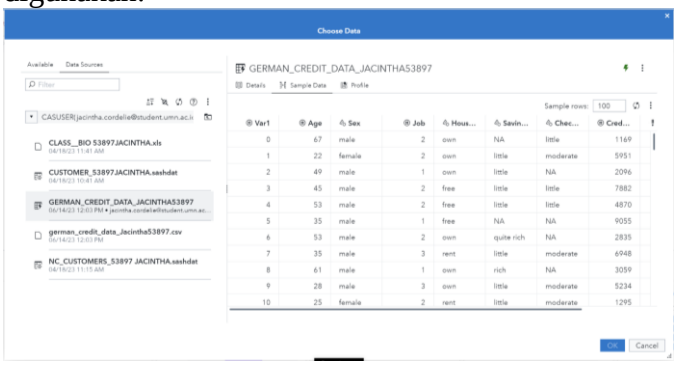
Setelah itu, kik "Upload"

4. Ketika *file* sudah berhasil ter-*upload*, maka tampilannya akan seperti dibawah ini, dimana *file* GERMAN_CREDIT_DATA_JACINTHA5 3897 sudah berada di dalam *folder* casuser.

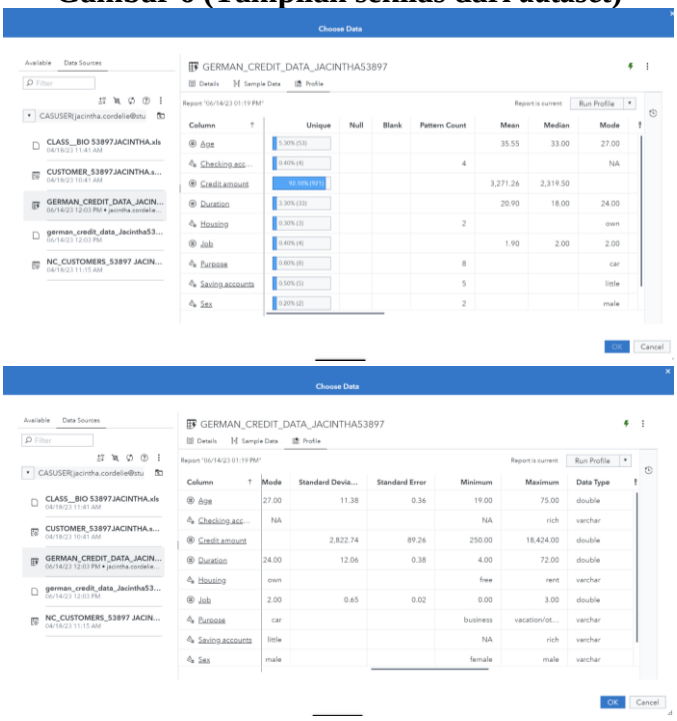


Gambar 5 (Tampilan *file* sudah ter-*uplaod*)

Berikut tampilan dan statistik dari *dataset* yang digunakan:



Gambar 6 (Tampilan sekilas dari *dataset*)



Gambar 7 (Tampilan statistik dari *dataset*)

Dari hasilnya dapat dilihat bahwa setiap atribut tidak memiliki data yang *null* dan *blank*. Lalu, hanya atribut yang dideteksi oleh SAS memiliki tipe data non-numerik yang memiliki nilai *pattern count*. Atribut yang memiliki *pattern count* adalah sebagai berikut:

Tabel 1 (*Pattern Count* setiap atribut non-numerik)

Checking account	4
Housing	2
Purpose	8
Saving accounts	5
Sex	2

Kemudian ada nilai *mean*, *median*, *standard deviation*, dan *standard error* yang hanya dimiliki atribut dideteksi oleh SAS memiliki tipe data numerik. Berikut list atributnya:

Tabel 2 (*Pattern Count* setiap atribut non-numerik)

Atribut	***Mean***	***Median***	***Standard Deviation***	***Standard Error***
Age	35,55	33	11,38	0,36
Credit amount	3.271,26	2.319,5	2.822,74	89,26
Duration	20,9	18	12,06	0,38
Job	1,9	2	0,65	0,02

Var1 tidak dianggap karena hanya atribut pembeda

Dan juga, ada nilai unique, mode, minimum, maximum yang dimiliki oleh hampir seluruh atribut:

Tabel 3 (*Pattern Count* setiap atribut non-numerik)

atribut	***unique***	***mode***	***minimum***	***maximum***
Age	5,3%	27	19	75
Checking account	0,4%	NA	NA	rich
Credit amount	92,1%	*	250	18.424
Duration	3,3%	24	4	72
Housing	0,3%	own	free	rent
Job	0,4%	2	0	3
Purpose	0,8%	car	business	vacation/others

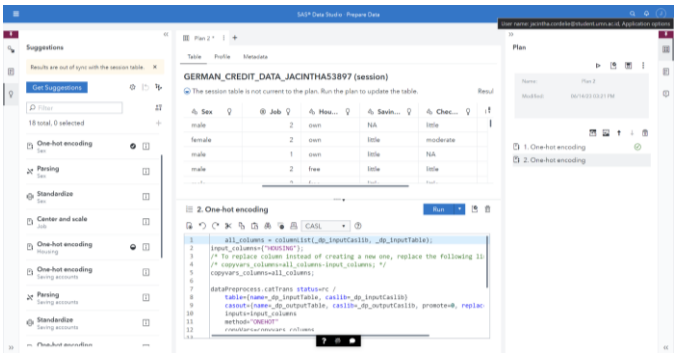
Saving accounts	0,5%	little	NA	rich
Sex	0,2%	male	female	male

*Atribut Credit amount tidak memiliki nilai *mode*, karena persentase nilai *unique* yang tinggi

C. Data Preparation

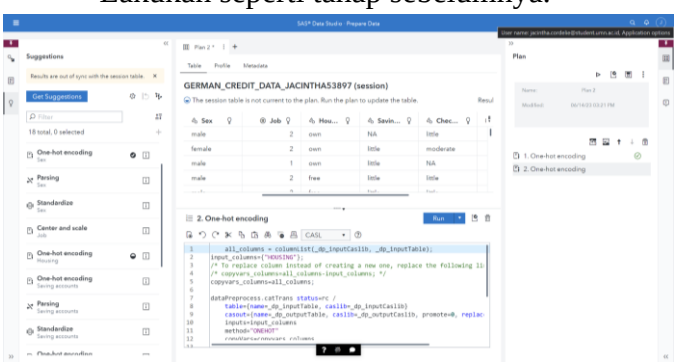
Pada tahap ini, peneliti akan melakukan preparasi data sebelum dibuat model. Berikut step yang dilakukan peneliti untuk melakukan tahap ini:

1. Pada *bar* sebelah kanan, klik logo lampu bohlam. Ketika tab "Suggestion", klik tombol "Get Suggestion" dan muncul 18 saran untuk diterapkan dalam data GERMAN_CREDIT_DATA_JACINTHA53897. Dan pada data ini, saya akan melakukan *one-hot encoding* pada semua atribut non-numerik. Pertama, peneliti akan melakukan *one-hot encoding* pada atribut Sex, dan kemudian pada atribut Housing. Klik kanan pada One-hot Encoding Housing dalam *tab* "Suggestions", dan pilih "Plan", dan plan untuk melakukan *one-hot encoding* akan muncul pada *tab* "Plan" disebelah kanan.



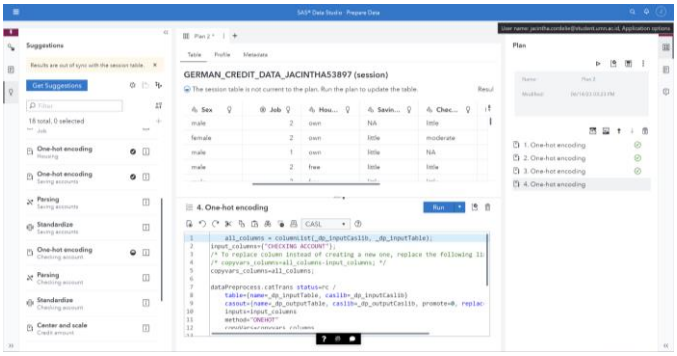
Gambar 8 (Tampilan setelah melakukan *one-hot encoding* pada atribut Sex, sebelum atribut Housing)

2. Selanjutnya, peneliti akan melakukan *one-hot encoding* pada atribut Saving accounts. Lakukan seperti tahap sebelumnya.



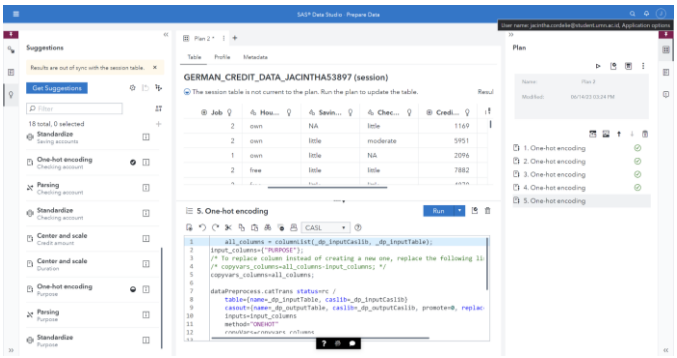
Gambar 9 (Tampilan sebelum melakukan *one-hot encoding* pada atribut Saving accounts)

3. Kemudian, peneliti akan melakukan one-hot encoding pada atribut Checking account.



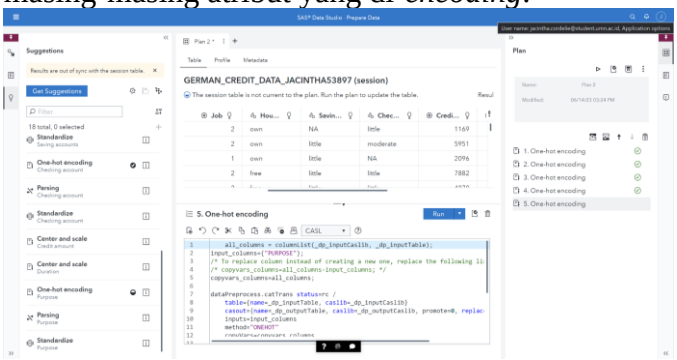
Gambar 10 (Tampilan sebelum melakukan *one-hot encoding* pada atribut Checking account)

4. Kemudian, peneliti akan melakukan one-hot encoding pada atribut Purpose.



Gambar 11 (Tampilan sebelum melakukan *one-hot encoding* pada atribut Purpose)

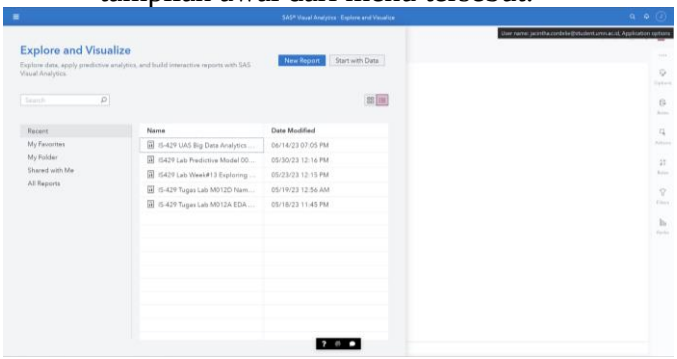
Hasil dari *one-hot encoding* pada kelima atribut non-numerik, dimana akan terbuat kolom baru untuk masing-masing atribut yang di-*encoding*.



Gambar 12 (Tampilan *dataset* setelah dilakukan *one-hot encoding*)

Berikutnya, peneliti akan mengubah semua value 'NA' dalam atribut Saving accounts dan Checking account. Tahap-tahap yang dilakukan peneliti adalah sebagai berikut:

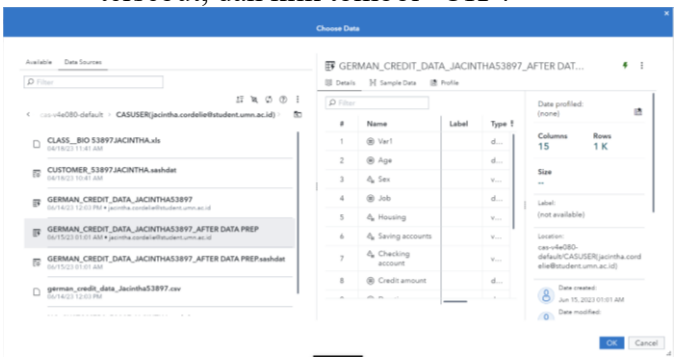
1. Pertama-tama, peneliti akan beralih ke menu Explore and Visualize. Berikut merupakan tampilan awal dari menu tersebut.



Gambar 13 (Tampilan menu Explore and Visualize)

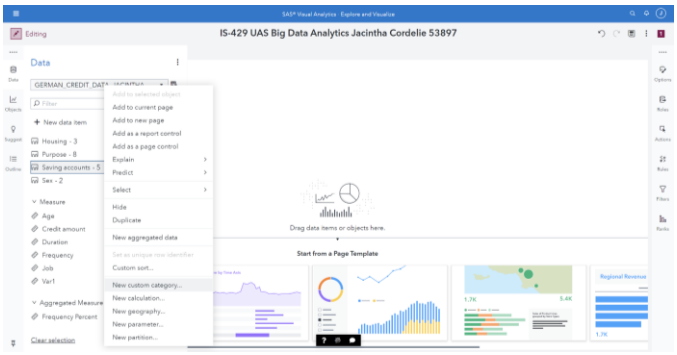
Pada tampilan ini, klik tombol "Start with Data" untuk memilih data yang digunakan untuk membuat report visualisasi.

2. Selanjutnya, akan muncul *pop-up* "Choose Data". Dalam *pop-up* ini, pilih menu "Data Source" → cas-v4080-default → CASUSER([e-mail]) dan memilih data yang sebelumnya sudah dibuat, yaitu GERMAN_CREDIT_DATA_JACINTHA53897_AFTER DATA PREP.sashdat. Namun, karena *file* .sashdat tidak dapat digunakan, peneliti perlu melakukan *load data* terlebih dahulu. Hasil dari *load data* ini merupakan *file* GERMAN_CREDIT_DATA_JACINTHA53897_AFTER DATA PREP. Pilih *file* tersebut, dan klik tombol "OK".



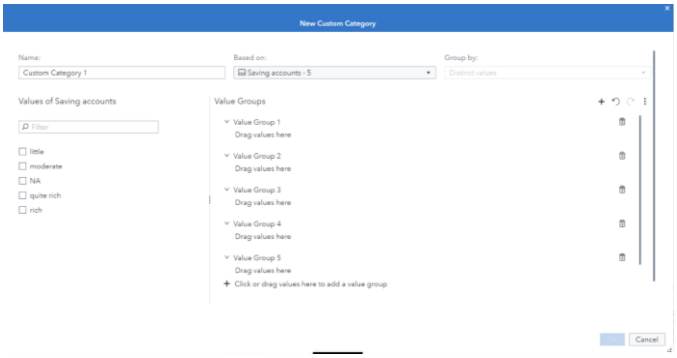
Gambar 14 (Tampilan *pop-up* Choose Data)

3. Berikutnya, peneliti akan melakukan penggantian *value* 'NA' pada atribut Saving accounts. peneliti buka *tab* "Data" yang terletak di *bar* sebelah kanan layar, lalu klik kanan pada atribut Saving accounts, dan pilih "New custom category"



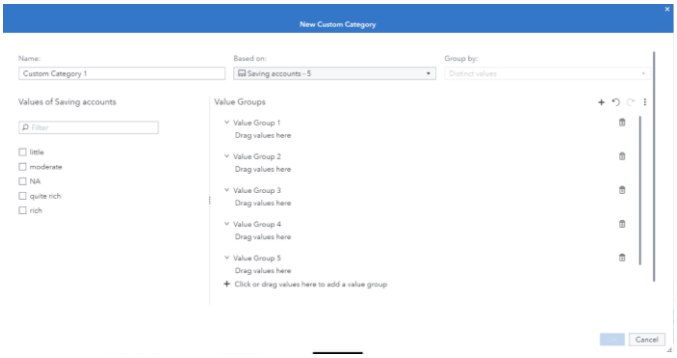
Gambar 15 (Tampilan klik kanan pada atribut Saving accounts)

4. Lalu, akan muncul *pop-up* "New Custom Category", dan peneliti akan memencet tombol + (pada kotak merah) sebanyak 4 kali, agar sesuai dengan jumlah *value* yang berada di sebelah kiri.



Gambar 16 (Tampilan *pop-up* New Custom Category)

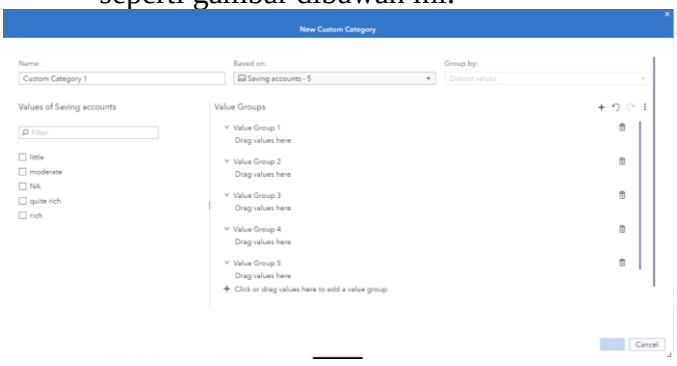
5. Berikutnya, peneliti akan *drag and drop* setiap *value* yang ada di sebelah kiri ke area yang bertuliskan "Drag values here" dibawah tulisan "Value Group 1", dst. *Drag* masing-masing *value* ke setiap "Value Group" yang ada.



Gambar 17 (Tampilan hasil *drag and drop*)

6. Kemudian, peneliti akan mengubah kelima "Value Group" yang ada dengan *value* yang sudah di *drag and drop* tadi, kecuali "Value Group 3" yang akan diubah dengan "not mentioned". Tidak lupa juga untuk mengganti "Name" menjadi "Saving

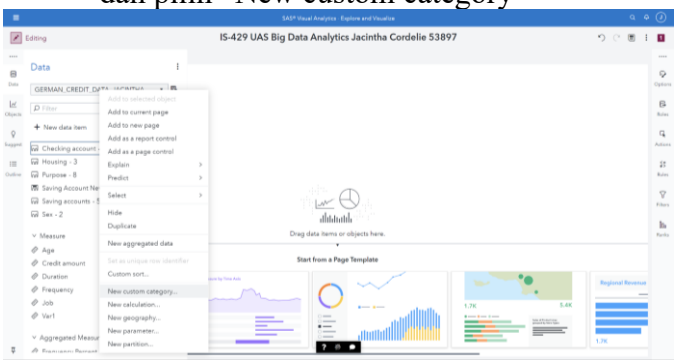
Account New" yang akan menjadi nama atribut baru yang akan dibuat. Hasilnya akan seperti gambar dibawah ini.



Gambar 18 (Tampilan setelah mengganti "Group Value" dan mengganti nama atribut)

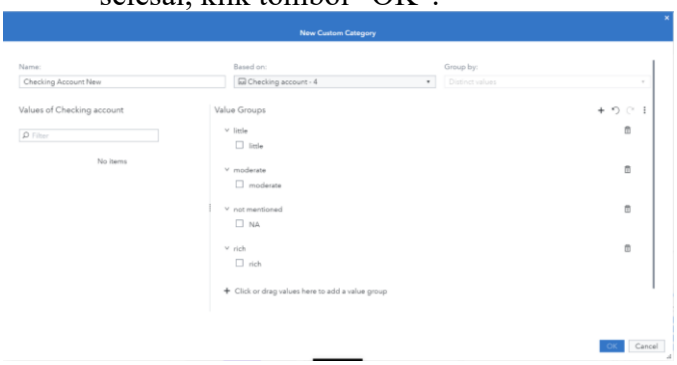
Setelah selesai, klik tombol "OK".

7. Dapat dilihat pada kotak merah dibawah bahwa atribut "Saving Account New" telah dibuat. Selanjutnya, peneliti akan melakukan hal yang sama pada atribut Checking account. Klik kanan pada atribut tersebut, dan pilih "New custom category"



Gambar 19 (Tampilan klik kanan pada atribut Checking accounts)

8. Lakukan *step* yang sama seperti yang sebelumnya. Nama atribut baru kali ini adalah "Checking Account New". Setelah selesai, klik tombol "OK".



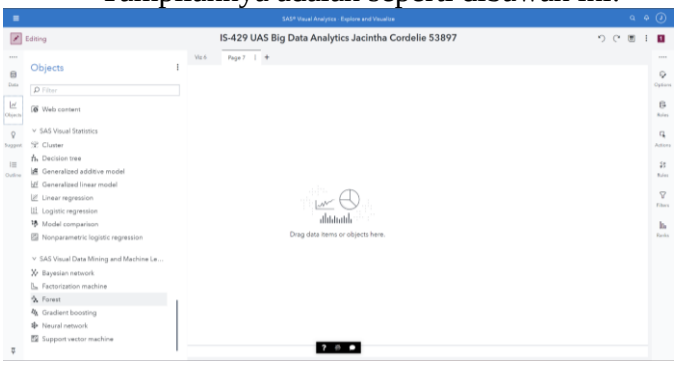
Gambar 20 (Tampilan setelah mengganti "Group Value" dan mengganti nama atribut)

D. Modeling

Pada tahap ini, peneliti akan membuat model prediksi. Peneliti menggunakan 3 algoritma yang berbeda untuk membuat model, yaitu Algoritma Forest, SVM, dan Decision Tree. Berikut tahap-tahap yang dilakukan peneliti untuk membuat model tersebut:

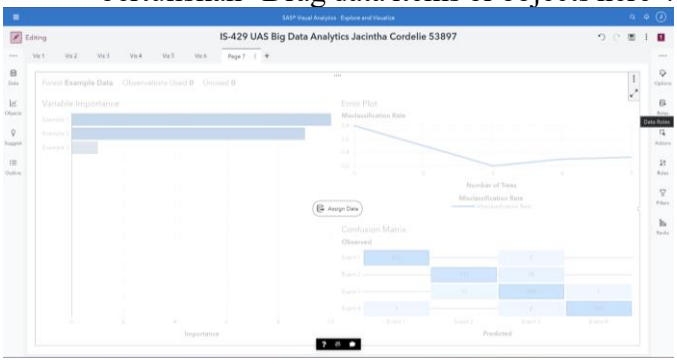
Forest

1. Pertama-tama, peneliti akan masuk ke menu Explore and Visualize, lalu peneliti kan meng-klik *tab* "Objects" di sebelah kiri layar. Selanjutnya peneliti *scroll down* dan mencari "Forest" pada kategori visualisasi "SAS Visual Data Mining and Machine Learning". Tampilannya adalah seperti dibawah ini:



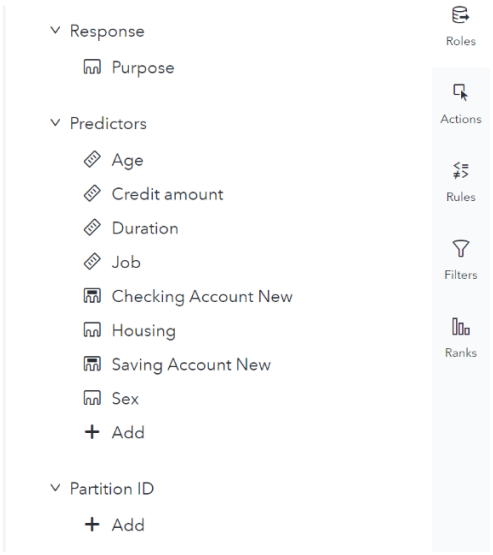
Gambar 21 (Tampilan "Forest" pada kategori visualisasi "SAS Visual Data Mining and Machine Learning")

2. Selanjutnya, peneliti akan melakukan *drag and drop* "Forest" ke layar utama yang bertuliskan "Drag data items or objects here".



Gambar 22 (Tampilan setelah *drag and drop* "Forest")

3. Kemudian, peneliti akan membuka *tab* "Roles" yang berada di sebelah kanan layar untuk memasukkan atribut apa saja yang akan digunakan untuk membuat model. Tampilan *tab* "Roles" yang sudah diisi adalah sebagai berikut:



Gambar 23 (Tampilan setelah *drag and drop* "Forest")

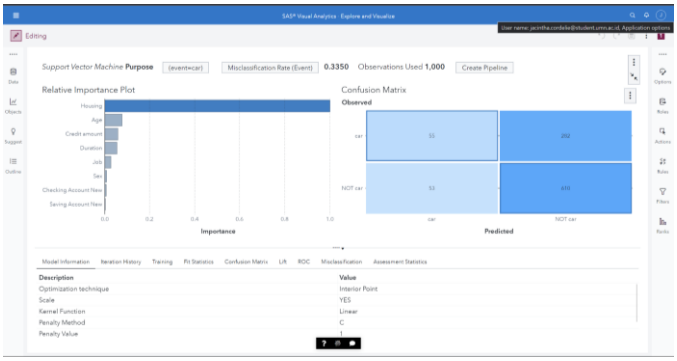
Saat data mulai dimasukkan, visualisasi akan mulai muncul.

4. Selanjutnya, peneliti akan mengganti *value* KS (Youden) ke Misclassification Rate, dan hasil akhir akan menjadi seperti dibawah ini:

Gambar 24 (Tampilan hasil akhir model dengan Algoritma Forest)

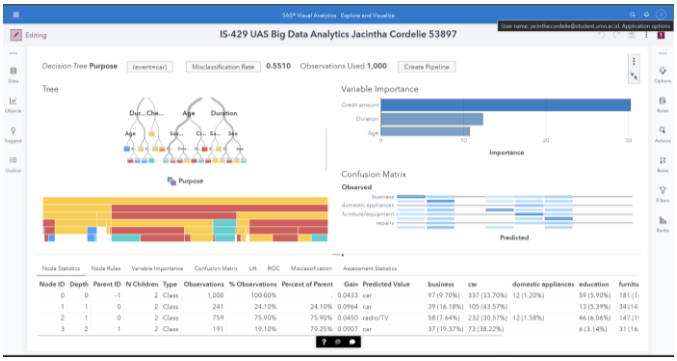
Proses pembuatan model dengan algoritma lain juga melewati tahap yang serupa. Berikut hasil akhir dari setiap model:

SVM



Gambar 25 (Tampilan hasil akhir model dengan Algoritma SVM)

Decision Tree



Gambar 26 (Tampilan hasil akhir model dengan Algoritma Decision Tree)

E. Evaluation

Pada tahap ini, peneliti akan mengevaluasi setiap model yang telah dibuat. Berikut evaluasinya:

Forest

1. *Confusion Matrix*

   Confusion matrix ini memiliki fungsi untuk mengetahui berapa prediksi yang benar. Berikut tampilannya:

Gambar 27 (Tampilan *Confusion Matrix* Model Forest)

Dapat dilihat dari *confusion matrix* bahwa jumlah prediksi pada setiap tujuan peminjaman adalah sebagai berikut:

Tabel 4 (*Insight* dari *Confusion Matrix* Model Forest)

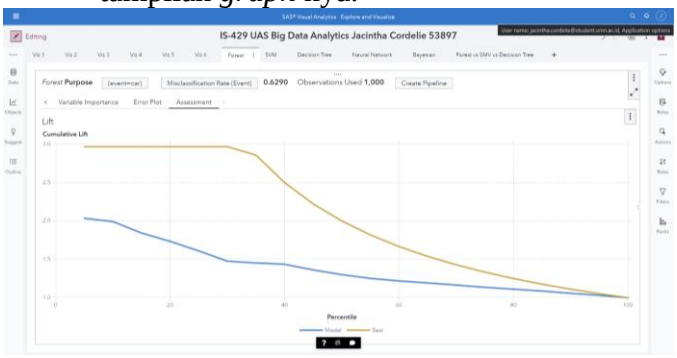
Tujuan Peminjaman	Jumlah Prediksi Benar
business	1
car	332
domestic appliance	0
education	0
furniture/equipment	9
radio/TV	9

repairs	0
vacation/others	0

Dapat dilihat dari *table* diatas, bahwa prediksi satu-satunya tujuan peminjaman yang memiliki jumlah prediksi benar diatas 10 adalah car. Sisanya kurang dari 10, seperti business yang hanya 1 dan furniture/equipment dan radio/TV yang hanya 9. Model yang dihasilkan memiliki ketimpangan, dimana sebagian besar hasil prediksinya adalah car. Hal ini bisa saja disebabkan oleh jumlah data yang cukup timpang, dimana model mempelajari secara detail ciri-ciri yang mengarah ke pinjaman untuk tujuan mobil daripada untuk tujuan lain.

2. *Lift Graph*

Graph ini berfungsi untuk mengevaluasi kinerja model yang telah dibuat. Graph ini akan melakukan evaluasi dalam memprediksi tujuan peminjaman yang untuk keperluan mobil (car), atau bukan. Berikut tampilan *graph* nya:



Gambar 28 (Tampilan *Lift Graph* Model Forest)

Dari hasil visualisasi diatas, dapat disimpulkan bahwa:

- 5% pertama hingga keenam menghasilkan nilai performa terbaik (*best*) yang tertinggi, yaitu 2,9674. Diantara keenam 5% data ini, yang keenam memiliki jarak terjauh antara nilai performa model (*model*) dan nilai performa terbaiknya, yaitu 0,8054 dan 2,9674.
- Nilai *model* semakin menurun hingga datanya pada 30%, dan selebihnya naik turun tidak menentu.
- Nilai *best* menurun drastis dari 2,1985 ke 0 pada presentasi data 35% ke 40%.
- Nilai *cumulative model* dan *cumulative best* semakin menurun seiring bertambahnya persentase data.

3. *ROC Graph*

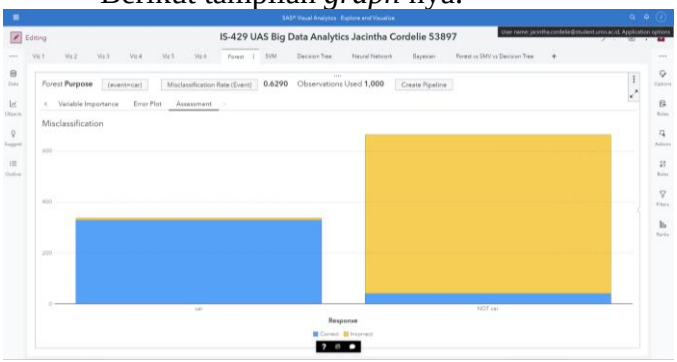
Graph ini berfungsi untuk mengevaluasi kinerja model yang telah dibuat. Graph ini akan melakukan evaluasi dalam memprediksi tujuan peminjaman yang untuk keperluan mobil (car), atau bukan. Berikut tampilan *graph* nya:

Gambar 29 (Tampilan *ROC Graph* Model Forest)

Dari visualisasi diatas, didapat nilai *sensitivity* sebesar 0,5757 dan nilai 1-*specificity* sebesar 0,312 dapat dianggap cukup baik.

4. *Misclassification Graph*

Graph ini berfungsi untuk mengetahui berapa banyak jumlah prediksi yang true positive, false positive, true negative dan false negative dalam memprediksi apakah pinjaman untuk keperluan mobil atau bukan. Berikut tampilan *graph* nya:



Gambar 30 (Tampilan *Misclassification Graph* Model Forest)

Berikut hasilnya jika diinterpretasikan dalam bentuk tabel:

Tabel 5 (*Insight* dari *Misclassification Graph* Model Forest)

False Positive	622	7	*False Negative*

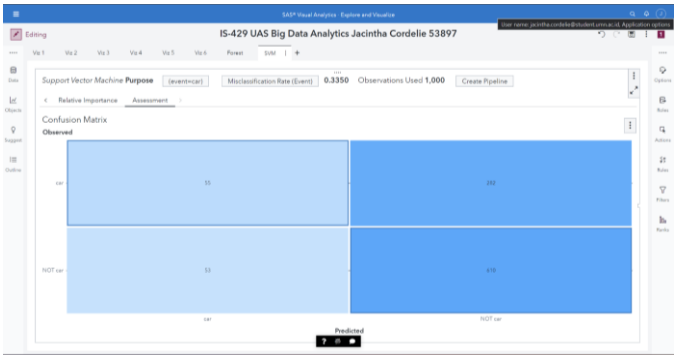
True Positive	330	41	*True Negative*

Dari hasilnya, dapat dilihat bahwa jumlah *true positive* yang lebih banyak daripada *false negative* berarti model sudah sangat baik untuk melakukan prediksi apakah tujuan peminjaman adalah untuk kepentingan mobil.

SVM

1. *Confusion Matrix*

   Berikut tampilannya *Confusion Matrix* dari Model SVM:

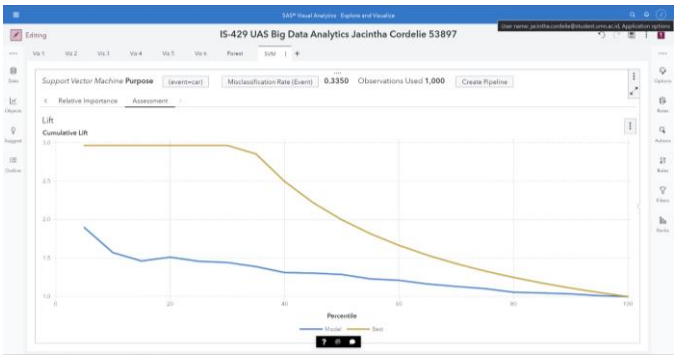


Gambar 31 (Tampilan *Confusion Matrix* Model SVM)

Dari hasil *confusion matrix*, dapat dilihat bahwa model ini bagus untuk memprediksi tujuan peminjaman yang bukan untuk keperluan mobil karena nilai *true negative* yang lebih besar daripada *true positive*.

2. *Lift Graph*

   Berikut tampilan *Lift Graph* dari Model SVM:



Gambar 32 (Tampilan *Lift Graph* Model SVM)

Dari hasil visualisasi diatas, dapat disimpulkan bahwa:

- 5% pertama hingga keenam menghasilkan nilai performa terbaik (*best*) yang tertinggi, yaitu 2,9674. Diantara keenam 5% data ini, yang ketiga memiliki jarak terjauh antara nilai performa model (*model*) dan nilai performa terbaiknya, yaitu 1,4639 dan 2,9674.
- Nilai *model* semakin menurun hingga datanya pada 15%, dan selebihnya naik turun tidak menentu.
- Nilai *best* menurun drastis dari 2,1985 ke 0 pada presentasi data 35% ke 40%.
- Nilai *cumulative model* terdapat sedikit kenaikan dan dilanjutkan dengan menurun, dan *cumulative best* semakin menurun seiring bertambahnya persentase data.

3. *ROC Graph*

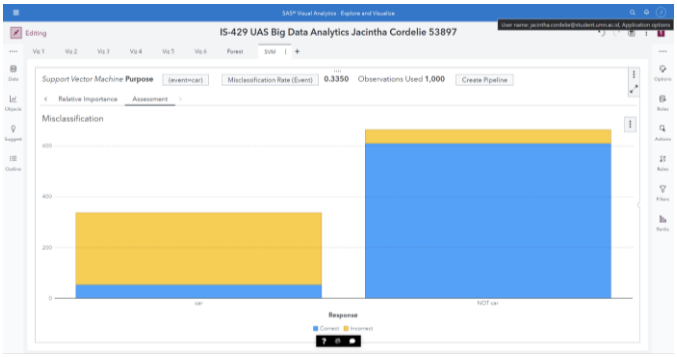
   Berikut tampilan *ROC Graph* dari Model SVM:

Gambar 33 (Tampilan *ROC Graph* Model SVM)

Dari visualisasi diatas, didapat nilai *sensitivity* sebesar 0,1632 dan nilai 1-*specificity* sebesar 0,080 dapat dianggap kurang baik, karena nilai *sensitivity* yang rendah dianggap kurang baik sedangkan nilai 1-*specificity* yang rendah dianggap baik.

4. *Misclassification Graph*

   Berikut tampilan *Misclassification Graph* dari Model SVM:



Gambar 34 (Tampilan *Misclassification Graph* Model SVM)

Berikut hasilnya jika diinterpretasikan dalam bentuk tabel:

Tabel 6 (*Insight* dari *Misclassification Graph* Model SVM)

False	53	282	*False*

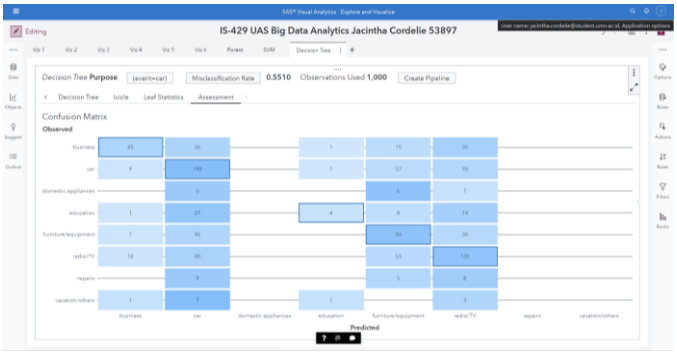
Positive			*Negative*
True Positive	55	610	*True Negative*

Dari hasilnya, dapat dilihat bahwa jumlah *true positive* yang lebih sedikit daripada *false negative* berarti model masih kurang baik untuk melakukan prediksi apakah tujuan peminjaman adalah untuk kepentingan mobil.

Decision Tree

1. *Confusion Matrix*

   Berikut tampilannya *Confusion Matrix* dari Model Decision Tree:



Gambar 35 (Tampilan *Confusion Matrix* Model Decision Tree)

Dapat dilihat dari *confusion matrix* bahwa jumlah prediksi pada setiap tujuan peminjaman adalah sebagai berikut:

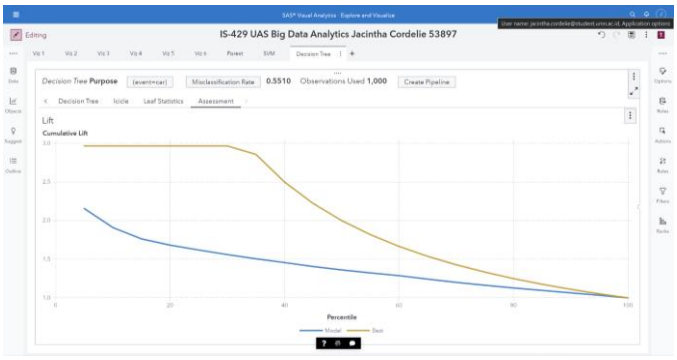
Tabel 7 (*Insight* dari *Confusion Matrix* Model Decision Tree)

Tujuan Peminjaman	Jumlah Prediksi Benar
business	25
car	192
domestic appliance	0
education	4
furniture/equipment	93
radio/TV	135
repairs	0
vacation/others	0

Dari tabel diatas, dapat dilihat bahwa masih ada beberapa yang hasil prediksi benarnya 0 seperti vacation/others, repairs, domestic appliance. Hal ini dapat disebabkan oleh jumlah data yang sangat sedikit dibandingkan tujuan peminjaman yang memiliki jumlah data banyak, seperti car, radio/TV, dan furniture/equipment.

2. *Lift Graph*

   Berikut tampilan *Lift Graph* dari Model Decision Tree:



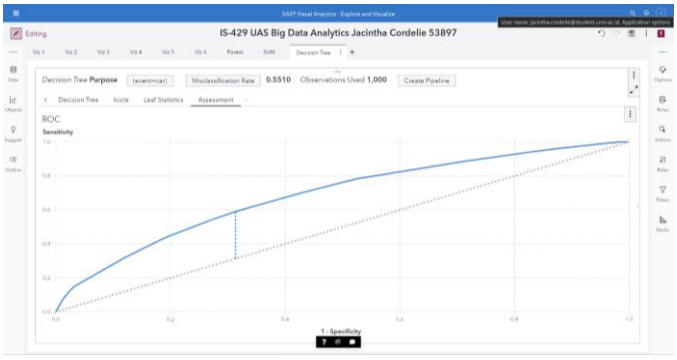
Gambar 36 (Tampilan *Lift Graph* Model Decision Tree)

Dari hasil visualisasi diatas, dapat disimpulkan bahwa:

- 5% pertama hingga keenam menghasilkan nilai performa terbaik (*best*) yang tertinggi, yaitu 2,9674. Diantara keenam 5% data ini, yang keenam memiliki jarak terjauh antara nilai performa model (*model*) dan nilai performa terbaiknya, yaitu 1,5603 dan 2,9674.
- Baik nilai *model*, *best*, *cumulative model* dan *cumulative best* semuanya semakin menurun seiring dengan semakin tinggi persentase datanya.

3. *ROC Graph*

   Berikut tampilan *ROC Graph* dari Model Decision Tree:

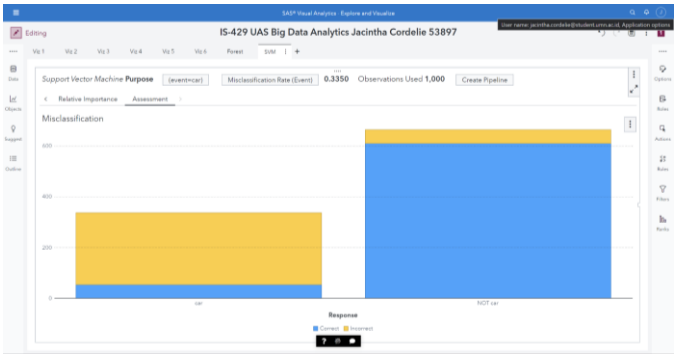


Gambar 37 (Tampilan *ROC Graph* Model Decision Tree)

Dari visualisasi diatas, didapat nilai *sensitivity* sebesar 0,5905 dan nilai 1-*specificity* sebesar 0,314 dapat dianggap cukup baik, karena semakin besar nilai *sensitivity* dan semakin kecil nilai 1-*specificity*, maka model dinilai semakin baik.

4. *Misclassification Graph*

Berikut tampilan *Misclassification Graph* dari Model Decision Tree:



Gambar 38 (Tampilan *Misclassification Graph* Model Decision Tree)

Berikut hasilnya jika diinterpretasikan dalam bentuk tabel:

Tabel 8 (*Insight* dari *Misclassification Graph* Model Decision Tree)

False Positive	21	287	*False Negative*
True Positive	50	642	*True Negative*

Dari hasilnya, dapat dilihat bahwa jumlah *true positive* yang lebih kecil daripada *false negative* berarti model masih kurang baik untuk melakukan prediksi apakah tujuan peminjaman adalah untuk kepentingan mobil.

F. Deployment

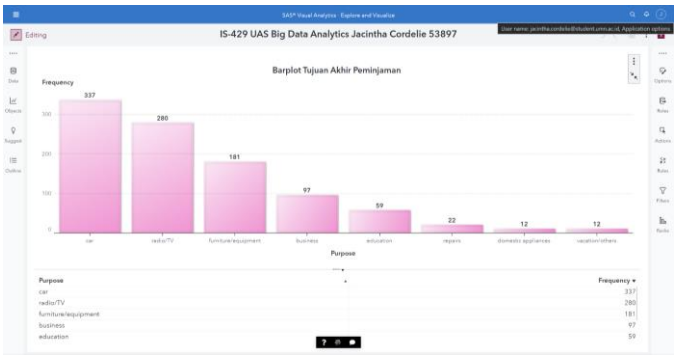
Tahap terakhir adalah tahap dimana peneliti akan menerapkan model yang sudah dibuat. Model yang terbaik sudah siap untuk digunakan secara real time, dimana model ini dapat digunakan untuk memprediksi peminjaman yang akan terjadi kedepannya.

IV. HASIL DAN PEMBAHASAN

Sebelum peneliti memasuki hasil dari setiap model, pertama-tama peneliti akan memaparkan visualisasi yang akan membantu dalam tahap hasil dan pembahasan.

Visualisasi 1 (*Barplot*)

Visualisasi yang pertama berupa *barplot* yang menunjukkan jumlah pinjaman untuk kedelapan tujuan pinjaman yang ada. Berikut visualisasinya:



Gambar 39 (Tampilan Visualisasi 1)

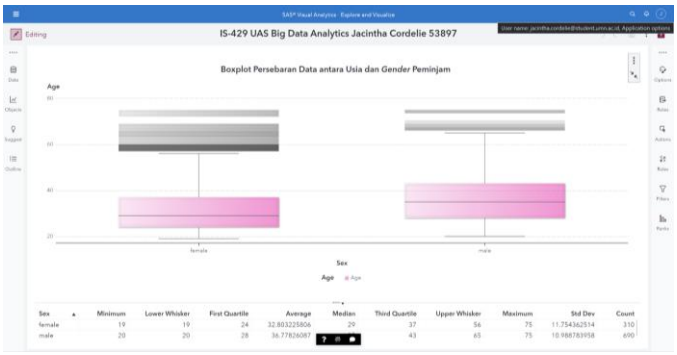
Dari visualisasi diatas, dapat disimpulkan urutan tujuan peminjaman yang paling banyak jumlah peminjamnya hingga paling sedikit yaitu:

Tabel 9 (*Insight* dari Visualisasi 1)

	Tujuan Peminjaman	Keterangan	Jumlah Peminjaman
1	car	keperluan berkaitan dengan mobil	337
2	radio/TV	keperluan berkaitan dengan radio atau TV	280
3	furniture/equipment	keperluan berkaitan dengan furnitur	181
4	business	keperluan berkaitan dengan bisnis	97
5	education	keperluan berkaitan dengan edukasi	59
6	repairs	keperluan perbaikan	22
7	domestic appliances	keperluan berkaitan dengan peralatan rumah tangga	12
8	vacation/others	keperluan berkaitan dengan liburan	12

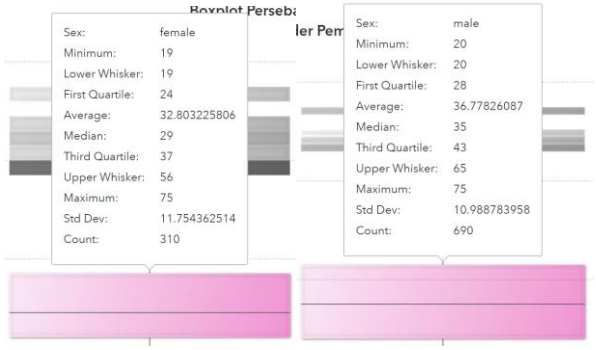
Visualisasi 2 (*Boxplot*)

Visualisasi kedua berbentuk *boxplot*, dimana disini akan diperlihatkan bagaimana distribusi jumlah pinjaman berdasarkan usia peminjam (Age) dan gender peminjam (Sex). Berikut visualisasinya:



Gambar 40 (Tampilan Visualisasi 2)

Dari visualisasi diatas, dapat dilihat bahwa data yang berada dalam *boxplot* diberi warna *pink*, dan *outlier* diberi warna abu-abu. Semakin pekat warna abu-abunya, maka semakin banyak jumlah data *outlier* yang berada pada range tersebut. Dan dapat disimpulkan juga bahwa data *boxplot* pada *gender female* memiliki *outlier* yang lebih banyak daripada *male*. Berikut detail dari masing masing *boxplot*:



Gambar 41 (*Detail* Tampilan Visualisasi 2)

Dari detail masing-masing boxplot, dapat dilihat bahwa:

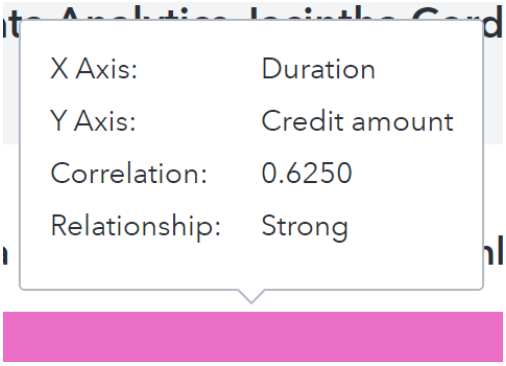
1. Kedua *boxplot* memiliki nilai *minimum* dan *lower whisker* yang sama, sedangkan nilai *maximum* dan *upper whisker* tidak sama. Hal ini dikarenakan data *outlier* hanya terdapat pada bagian atas *boxplot*.
2. Rata-rata usia peminjam ber-*gender female* lebih kecil daripada *male*. Hal ini menandakan bahwa secara keseluruhan, kebanyakan peminjam *female* memiliki usia yang lebih rendah daripada peminjam *male*.
3. Peminjam *female* memiliki nilai *standard deviation* yang lebih kecil daripada peminjam *male*. Hal ini menandakan bahwa variasi usia peminjam *female* tidak lebih beragam daripada variasi usia peminjam *male*.

Visualisasi 3 (*Confusion Matrix*)

Visualisasi ketiga berupa *confusion matrix*, dimana visualisasi ini bertujuan untuk menggambarkan korelasi antara lama peminjaman (Duration) dan jumlah peminjaman (Credit amount). Berikut visualisasinya:

Gambar 42 (Tampilan Visualisasi 3)

Jenis korelasi digambarkan dengan seberapa pekat warna dari *box*-nya. Semakin pekat warnanya, maka semakin kuat korelasinya. Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *strong positive*. Berikut *detail* dari *box* tersebut:



Gambar 43 (*Detail* Tampilan Visualisasi 3)

Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah 0.625 dan *relationship*-nya adalah strong. Hal ini sesuai dengan apa yang sudah disebutkan sebelumnya, bahwa korelasinya adalah *strong positive*.

Visualisasi 4 (*Treemap*)

Visualisasi keempat berupa *treemap*, dimana visualisasi ini bertujuan untuk menggambarkan banyak jumlah peminjam berdasarkan kategori kondisi jumlah uangnya (Saving Account New). Berikut visualisasinya:

Gambar 44 (Tampilan Visualisasi 4)

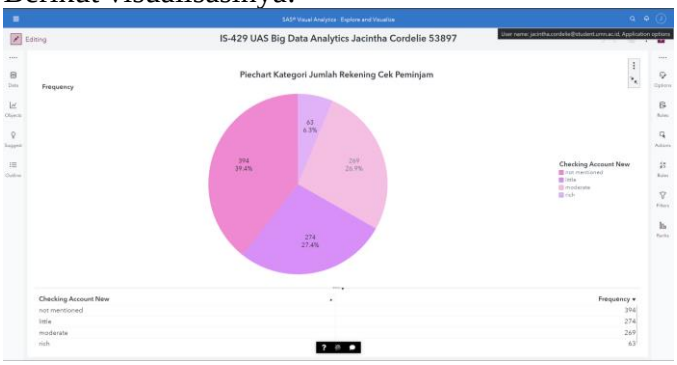
Dari visualisasi diatas, dapat disimpulkan bahwa urutan kategori kondisi jumlah uang yang paling banyak jumlah peminjamnya hingga paling sedikit yaitu:

Tabel 10 (*Insight* dari Visualisasi 4)

	Jumlah Uang	Keterangan	Jumlah Peminjaman
1	little	uang dalam tabungan berjumlah sedikit	603
2	not mentioned	jumlah uang dalam tabungan tidak disebutkan	183
3	moderate	uang dalam tabungan berjumlah sedang	103
4	quite rich	uang dalam tabungan berjumlah lumayan banyak	63
5	rich	uang dalam tabungan berjumlah banyak	48

Visualisasi 5 (*Pie Chart*)

Visualisasi kelima berupa *pie chart*, dimana visualisasi ini bertujuan untuk menggambarkan banyak jumlah peminjam berdasarkan kategori jumlah rekening ceknya (Checking Account New). Berikut visualisasinya:



Gambar 45 (Tampilan Visualisasi 5)

Dari visualisasi diatas, dapat disimpulkan bahwa urutan jumlah rekening cek yang paling banyak jumlah peminjamnya hingga paling sedikit yaitu:

Tabel 11 (*Insight* dari Visualisasi 5)

	Jumlah Uang	Keterangan	Jumlah Peminjaman
1	not mentioned	jumlah rekening cek tidak disebutkan	394
2	little	rekening cek berjumlah sedikit	274
3	moderate	rekening cek berjumlah sedang	269
4	rich	rekening cek berjumlah banyak	63

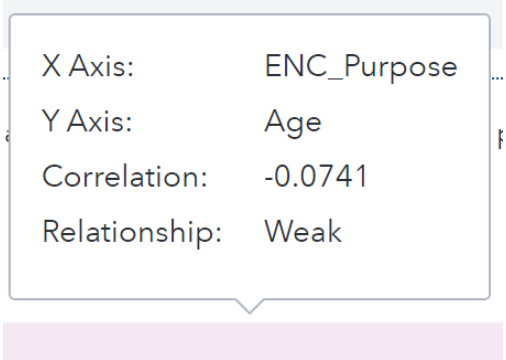
Visualisasi 6 (*Confusion Matrix*)

Visualisasi keenam berupa *confusion matrix*, dimana visualisasi ini bertujuan untuk menggambarkan korelasi antara tujuan peminjaman (ENC_Purpose) dan masing-masing atribut lainnya. Berikut visualisasinya:

1. Tujuan Peminjaman (ENC_Purpose) dan Usia Peminjam (Age)

Gambar 46 (Tampilan *Confusion Matrix* ENC_Purpose dan Age)

Sama seperti visualisasi ketiga, jenis korelasi digambarkan dengan seberapa pekat warna dari *box*-nya. Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *weak negative*. Berikut *detail* dari *box* tersebut:



Gambar 47 (*Detail Confusion Matrix* ENC_Purpose dan Age)

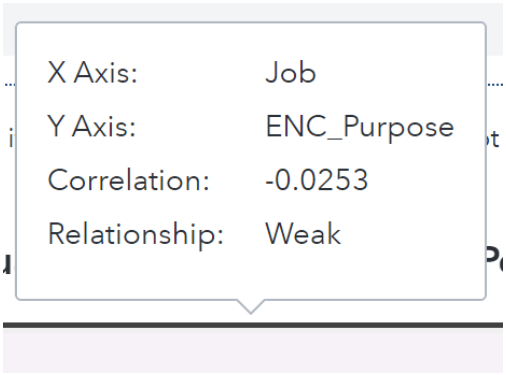
Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah 0.625 dan *relationship*-

nya adalah strong. Hal ini sesuai dengan apa yang sudah disebutkan sebelumnya, bahwa korelasinya adalah *strong positive*.

2. Tujuan Peminjaman (ENC_Purpose) dan Jenis Pekerjaan (Job)

Gambar 48 (Tampilan *Confusion Matrix* ENC_Purpose dan Job)

Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *weak negative*. Berikut *detail* dari *box* tersebut:



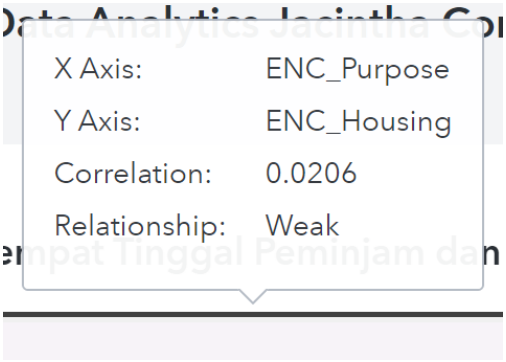
Gambar 49 (*Detail Confusion Matrix* ENC_Purpose dan Job)

Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah -0.0253 dan *relationship*-nya adalah *weak*. Hal ini sesuai dengan apa yang sudah disebutkan sebelumnya, bahwa korelasinya adalah *weak negative*.

3. Tujuan Peminjaman (ENC_Purpose) dan Status Tempat Tinggal (ENC_Housing)

Gambar 50 (Tampilan *Confusion Matrix* ENC_Purpose dan ENC_Housing)

Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *weak negative*. Berikut detail dari box tersebut:



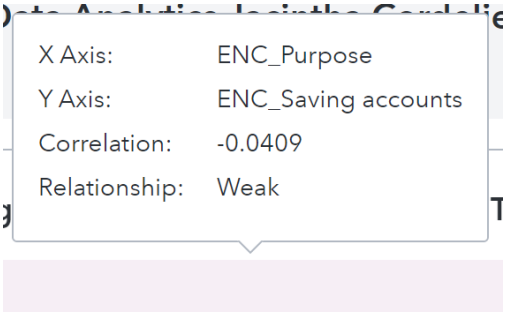
Gambar 51(*Detail Confusion Matrix* ENC_Purpose dan ENC_Housing)

Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah 0.0206 dan relationship-nya adalah *weak*. Hal ini tidak sesuai dengan apa yang sudah disebutkan sebelumnya, karena korelasinya adalah *weak positive*.

4. Tujuan Peminjaman (ENC_Purpose) dan Jumlah Uang dalam Tabungan Peminjam (ENC_Saving accounts)

Gambar 52 (Tampilan *Confusion Matrix* ENC_Purpose dan ENC_Saving accounts)

Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *weak negative*. Berikut detail dari box tersebut:



Gambar 53 (*Detail Confusion Matrix* ENC_Purpose dan ENC_Saving accounts)

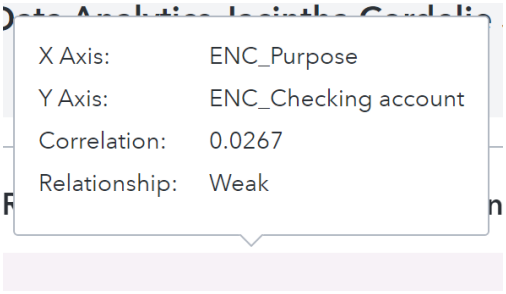
Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah -0.0409 dan relationship-nya adalah *weak*. Hal ini sudah sesuai dengan

apa yang sudah disebutkan sebelumnya, bahwa korelasinya adalah *weak negative*.

5. Tujuan Peminjaman (ENC_Purpose) dan Jumlah Rekening Cek Peminjam (ENC_Checking account)

Gambar 54 (Tampilan *Confusion Matrix* ENC_Purpose dan ENC_Checking account)

Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *weak negative*. Berikut detail dari box tersebut:



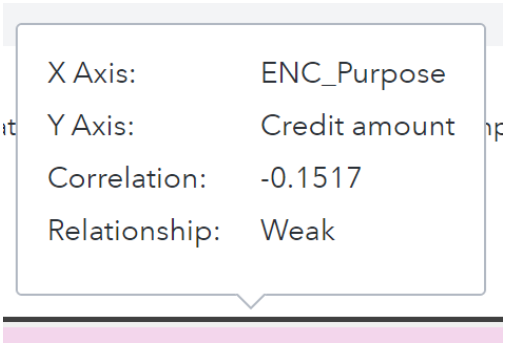
Gambar 55 (*Detail Confusion Matrix* ENC_Purpose dan ENC_Checking account)

Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah 0.0267 dan *relationship*-nya adalah *weak*. Hal ini tidak sesuai dengan apa yang sudah disebutkan sebelumnya, karena korelasinya adalah *weak positive*.

6. Tujuan Peminjaman (ENC_Purpose) dan Jumlah Peminjaman (Credit Amount)

Gambar 56 (Tampilan *Confusion Matrix* ENC_Purpose dan Credit Amount)

Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *weak positive*. Berikut detail dari box tersebut:



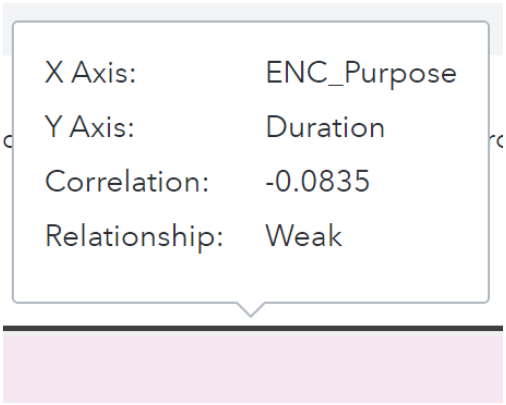
Gambar 57 (*Detail Confusion Matrix* ENC_Purpose dan Credit Amount)

Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah -0.1517 dan relationship-nya adalah *weak*. Hal ini tidak sesuai dengan apa yang sudah disebutkan sebelumnya, karena korelasinya adalah *weak negative*.

7. Tujuan Peminjaman (ENC_Purpose) dan Lama Peminjaman (Duration)

Gambar 58 (Tampilan *Confusion Matrix* ENC_Purpose dan Duration)

Dari warnanya, diambil kesimpulan bahwa korelasinya adalah *weak positive*. Berikut detail dari box tersebut:



Gambar 59 (*Detail Confusion Matrix* ENC_Purpose dan Duration)

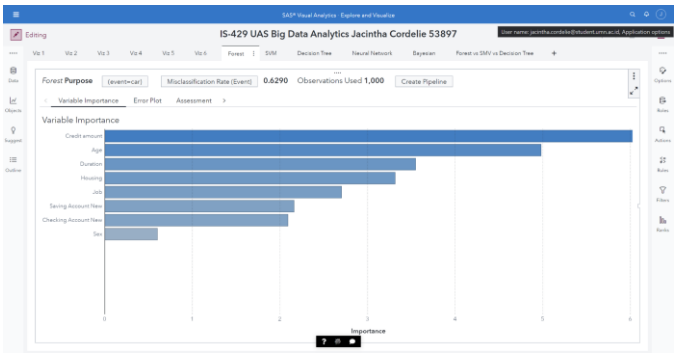
Dari *detail* diatas, dapat dilihat bawa nilai korelasinya adalah -0.0835 dan *relationship*-nya adalah *weak*. Hal ini tidak sesuai dengan apa yang sudah disebutkan sebelumnya, karena korelasinya adalah *weak negative*.

Selanjutnya, peneliti akan menjelaskan beberapa visualisasi yang dihasilkan oleh model.

Forest

1. *Variable Importance Graph*

   Graph ini memiliki fungsi untuk mengetahui seberapa besar peran masing-masing atribut yang termasuk dalam bagian predictor. Berikut tampilan graph nya:



Gambar 60 (Tampilan *Variable Importance Graph* Model Forest)

Dari visualisasi diatas, dapat disimpulkan atribut yang memiliki kontribusi paling besar hingga paling kecil yaitu:

Tabel 12 (*Insight* dari *Variable Importance Graph* Model Forest)

	Atribut	*Importance*
1	Credit amount	6,4686
2	Age	4,9819
3	Duration	3,5527
4	Housing	3,3159
5	Job	2,7030
6	Saving Account New	2,1597
7	Checking Account New	2,0936
8	Sex	2,4921

2. *Error Plot*

   Plot ini memiliki fungsi untuk mengetahui nilai *misclassification rate* dan nilai *misclassification rate out-of-bag* untuk setiap jumlah *tree* yang ada dalam model Forest. Berikut tampilan *graph* nya:

Gambar 61 (Tampilan *Error Plot* Model Forest)

Dari *plot* diatas, jika dilihat secara keseluruhan, dapat disimpulkan bahwa rata-rata nilai *misclassification rate* lebih rendah daripada nilai *misclassification rate out-of-bag*. Nilai *misclassification rate* terendah terdapat pada model dengan jumlah *tree*-nya 7, yaitu sebesar 0,6120, sedangkan nilai *misclassification rate out-of-bag* terendah terdapat pada model dengan jumlah *tree*-nya 1, yaitu sebesar 0,6532. Karena perbedaan nilai *misclassification rate* pada model dengan jumlah *tree*-nya 7 (0,6120) dan yang jumlah *tree*-nya 1 (0,6520) lebih besar daripada perbedaan nilai *misclassification rate out-of-bag* pada model dengan jumlah *tree*-nya 1 (0,6532) dan yang jumlah *tree*-nya 7 (0,6663), maka menggunakan jumlah *tree* yang dipilih adalah 1.

3. *Partial Dependence Plot (Credit amount)*

   Plot ini memiliki fungsi untuk mengetahui bagaimana setiap *value* dalam atribut Credit amount dalam mempengaruhi hasil prediksi dalam model. Berikut tampilan *graph* nya:

Gambar 62 (Tampilan *Partial Dependence Plot* Model Forest)

Dari visualisasi diatas, dapat disimpulkan bahwa pengaruh setiap *value* pada atribut Credit amount terhadap prediksi untuk tujuan pinjaman untuk keperluan mobil adalah sebagai berikut:

- Pada nilai Credit amount 704,53 hingga 1.613,05, nilai *predicted probability* nya naik secara drastis.
- Pada nilai Credit amount 1.613,05 hingga 5.247,85, nilai *predicted probability* nya naik namun tidak sedrastis yang sebelumnya.
- Pada nilai Credit amount 5.247,85 hingga 6.156,55, nilai *predicted probability* nya konstan. Tidak naik dan tidak turun.
- Pada nilai Credit amount 6.156,55 hingga 7.973,95, nilai *predicted probability* nya naik lagi secara drastis.
- Pada nilai Credit amount 7.973,95 hingga 8.882,65, nilai *predicted probability* nya menurun, namun tidak drastis.
- Pada nilai Credit amount 8.882,65 hingga 9.791,35, nilai *predicted probability* nya konstan. Tidak naik dan tidak turun.
- Pada nilai Credit amount 9.791,35 hingga 10.700,05, nilai *predicted probability* nya menurun secara drastis.
- Pada nilai Credit amount 10.700,05 hingga 11.608,75, nilai *predicted probability* nya konstan.
- Pada nilai Credit amount 11.608,75 hingga 12.517,45, nilai *predicted probability* nya menurun lagi secara drastis.
- Pada nilai Credit amount 12.517,45 hingga 17.969,65, nilai *predicted probability* nya konstan.

SVM

1. *Relative Importance Plot*

   Plot ini memiliki fungsi untuk mengetahui seberapa besar peran masing-masing atribut yang termasuk dalam bagian *predictor*. Berikut tampilan *graph* nya:

Gambar 63 (Tampilan *Relative Importance Plot* Model SVM)

Dari visualisasi diatas, dapat disimpulkan atribut yang memiliki kontribusi paling besar hingga paling kecil yaitu:

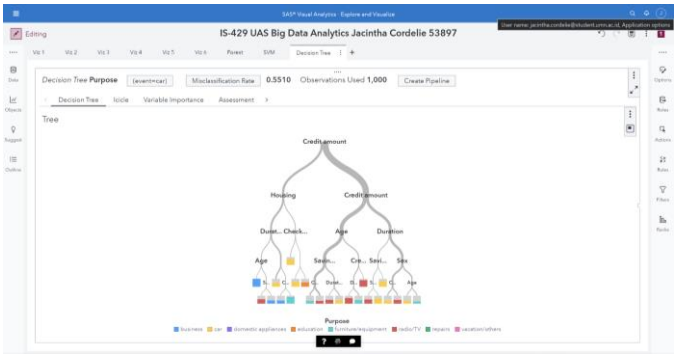
Tabel 13 (*Insight* dari *Relative Importance Plot* Model SVM)

	Atribut	*Importance*
1	Housing	1
2	Age	0,0774
3	Credit amount	0,0584
4	Duration	0,0558
5	Job	0,0274
6	Sex	0,0102
7	Checking Account New	0,0065
8	Saving Account New	0,0023

Decision Tree

1. *Decision Tree Graph*

   Pada *graph* ini, akan memperlihatkan hasil prediksi untuk semua tujuan peminjaman. Berikut merupakan *graph* nya:



Gambar 64 (Tampilan *Decision Tree Graph* Model Decision Tree)

Dari graph diatas, dapat dilihat bahwa hasil prediksi yang dilihat dari *leaf*-nya adalah sebagai berikut:

Tabel 14 (*Insight* dari *Decision Tree Graph* Model Decision Tree)

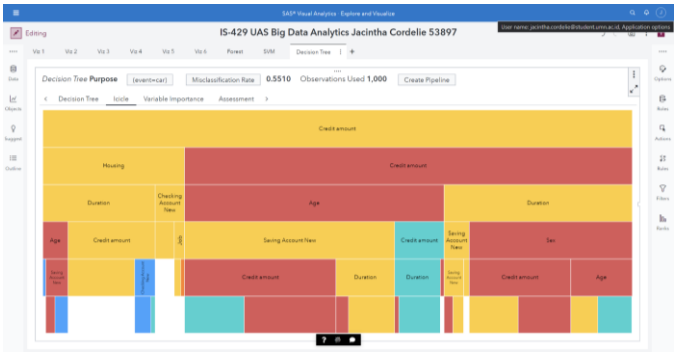
Warna	Tujuan Peminjaman	Jumlah *leaf*
kuning	car	8
merah	radio/TV	7

tosca	furniture/equipment	4
biru	business	3
oren	education	1

Dari tabel diatas, dapat dilihat bahwa car, radio/TV, dan furniture/equipment mendominasi sebagai hasil akhir prediksi. Hal ini bisa jadi disebabkan karena ketiganya memiliki jumlah pinjaman yang jauh lebih tinggi daripada tujuan peminjaman lainnya.

2. *Icicle Graph*

Graph ini berfungsi untuk memvisualisasikan kontribusi atribut-atribut pada pada *predictors* terhadap pembentukan struktur *decision tree*. Berikut tampilan *graph* nya:

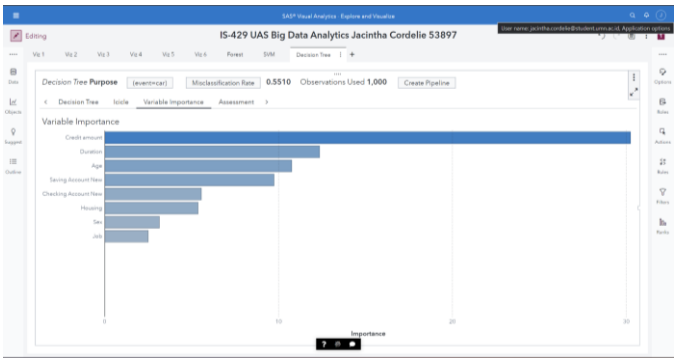


Gambar 65 (Tampilan *Icicle Graph* Model Decision Tree)

Dari graph diatas, dapat dilihat *root* dari *decision tree* merupakan atribut Credit amount, yang kemudian dilanjutkan dengan atribut Housing dan Credit Amount dan seterusnya. Warna pada setiap segi empat melambangkan *value* atribut Purpose yang memiliki jumlah *count* terbanyak.

3. *Variable Importance Graph*

Berikut tampilan *Variable Importance Graph* dari model Decision Tree:



Gambar 66 (Tampilan *Variable Importance Graph* Model Decision Tree)

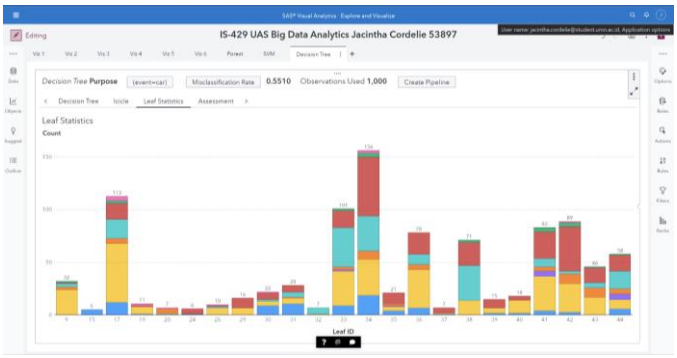
Dari visualisasi diatas, dapat disimpulkan atribut yang memiliki kontribusi paling besar hingga paling kecil yaitu:

Tabel 15 (*Insight* dari *Variable Importance Graph* Model Decision Tree)

	Atribut	*Importance*
1	Credit amount	30,2196
2	Duration	12,3518
3	Age	10,7455
4	Saving Account New	9,7276
5	Checking Account New	5,5615
6	Housing	5,3716
7	Sex	3,1573
8	Job	2,4921

4. *Leaf Statistic Graph*

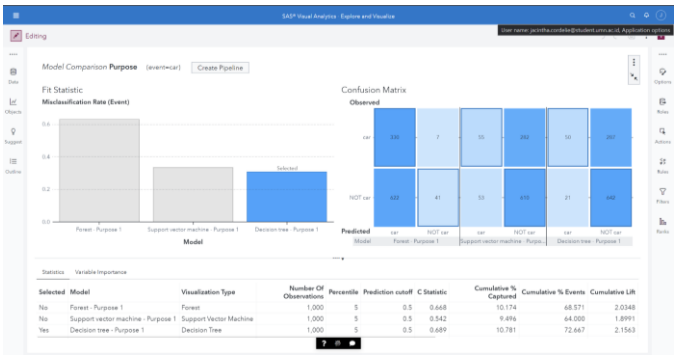
Graph ini memiliki fungsi untuk mengetahui distribusi setiap tujuan peminjaman di setiap *leaf* pada model *decision tree*. Sama seperti jumlah *leaf* pada *decision tree graph*, disini juga terdapat 23 *leaf*. Berikut tampilan *graph* nya:



Gambar 67 (Tampilan *Leaf Statistic Graph* Model Decision Tree)

Dapat dilihat bahwa warna merah, tosca, dan kuning memiliki jumlah *count* yang besar pada sebagian besar *leaf*.

Kemudian, peneliti akan membandingkan performa dari ketiga model tersebut. Berikut visualisasi keseluruhan hasil perbandingannya:

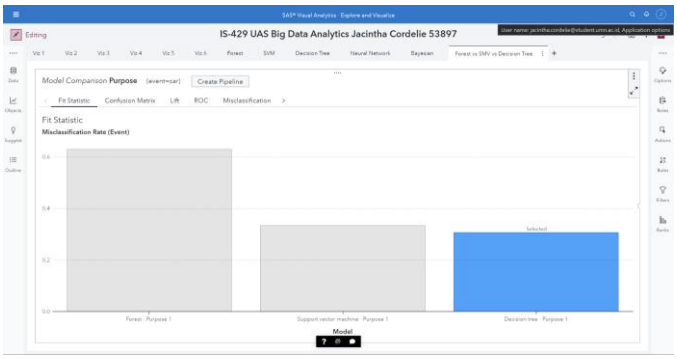


Gambar 68 (Tampilan Perbandingan Performa ketiga algoritma)

Berikut merupakan *detail chart* yang ada pada visualisasi Perbandingan model diatas:

1. *Fit Statistic*

   Statistic ini berfungsi untuk membandingkan *misclassification rate* dari ketiga model yang telah dibuat. Berikut *barplot*-nya:

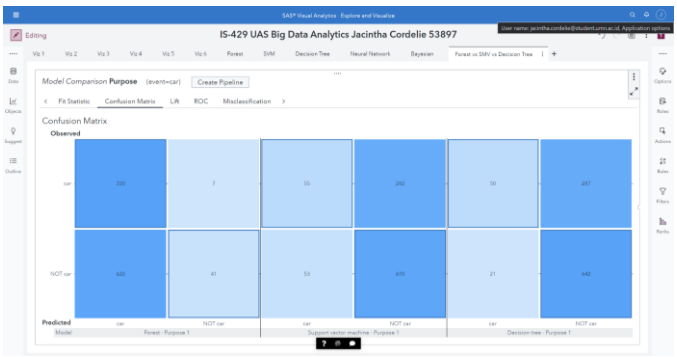


Gambar 69 (Tampilan *Fit Statistic* dari Ketiga Model)

Dapat dilihat bahwa model dengan *misclassification rate* terkecil adalah **Decision Tree**.

2. *Confusion Matrix*

   Confusion matrix ini memiliki fungsi untuk mengetahui perbandingan prediksi yang benar di antara ketiga model. Berikut tampilannya:



Gambar 70 (Tampilan *Confusion Matrix* Ketiga Model)

Dari ketiga *confusion matrix* di atas, dapat dilihat bahwa model yang paling baik untuk melakukan prediksi tujuan peminjaman untuk kepentingan mobil yang tepat adalah **Forest**.

3. *Cumulative Lift Graph*

   Graph ini berfungsi untuk mengevaluasi kinerja model yang telah dibuat. *Graph* ini akan melakukan evaluasi dalam memprediksi tujuan peminjaman yang untuk keperluan mobil (car), atau bukan. Berikut tampilan *graph* nya:

Gambar 71 (Tampilan *Cumulative Lift Graph* dari Ketiga Model)

Dari hasil visualisasi diatas, dapat disimpulkan bahwa:

Tabel 16 (*Insight* dari *Cumulative Lift Graph* Ketiga Model)

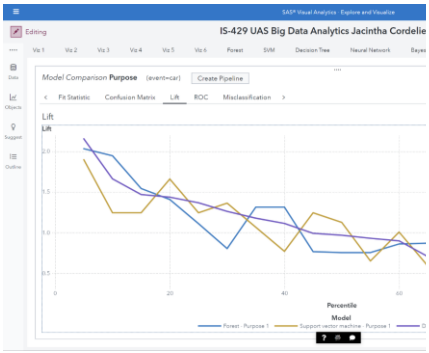
Persentase data	Model yang terbaik
5	Decision Tree
10	Forest
15	Forest
20	Forest
25	Decision Tree
30	Decision Tree
35	Decision Tree
40	Decision Tree
45	Decision Tree
50	Decision Tree
55	Decision Tree

60	Decision Tree
65	Decision Tree
70	Decision Tree
75	Decision Tree
80	Decision Tree
85	Decision Tree
90	Decision Tree
95	Decision Tree
100	Decision Tree, Forest, SVM

Pada saat data 100%, ketiganya memiliki nilai *cumulative lift* yang sama, yaitu 1. Dan juga dapat disimpulkan bahwa Model **Decision Tree** memiliki *cumulative lift* terbaik diantara ketiga model.

4. *Lift Graph*

   Berikut perbandingan *Lift Graph* untuk ketiga model:



Gambar 72 (Tampilan *Lift Graph* dari Ketiga Model)

Dari hasil visualisasi diatas, dapat disimpulkan bahwa:

Tabel 17 (*Insight* dari *Lift Graph* Ketiga Model)

Persentase data	Model yang terbaik
5	Decision Tree
10	Forest
15	Forest
20	SVM
25	Decision Tree
30	SVM
35	Forest
40	Forest
45	SVM
50	SVM
55	Decision Tree
60	SVM
65	Forest
70	Forest
75	SVM
80	Forest
85	SVM
90	SVM
95	SVM
100	SVM

Dapat disimpulkan bahwa Model **SVM** memiliki nilai *lift* terbaik di antara ketiga model.

5. *ROC Graph*

   Berikut perbandingan *ROC Graph* untuk ketiga model:

Gambar 73 (Tampilan *ROC Graph* dari Ketiga Model)

Dari *graph* diatas, dapat disimpulkan bahwa:

Tabel 17 (*Insight* dari *ROC Graph* Ketiga Model)

Model	*Sensitivity*	1-*specificity*
Decision Tree	0,5905	0,314
Forest	0,5757	0,312
SVM	0,1632	0,080

Dari *table* diatas, dapat disimpulkan bahwa model dengan ROC graph terbaik adalah model **Decision Tree**, karena model ini memiliki nilai *sensitivity* yang jauh lebih tinggi daripada model Forest. Dan juga, hal ini dikarenakan perbedaan nilai 1-*specificity* dari model Forest hanya lebih unggul sedikit daripada model Decision Tree, yaitu beda 0,002.

6. *Misclassification Graph*

Berikut perbandingan *Misclassification Graph* untuk ketiga model:

Gambar 74 (Tampilan *Misclassification Graph* dari Ketiga Model)

Dari ketiga *graph* diatas, dapat disimpulkan bahwa model **Forest** yang terbaik untuk memprediksi tujuan pinjaman untuk keperluan mobil, karena model ini memiliki nilai *true positif* yang tertinggi, yaitu 330.

V. KESIMPULAN

Dari hasil dan pembahasan sebelumnya, dapat disimpulkan bahwa Model Decision Tree adalah model yang terbaik untuk kasus ini. Pemilihan model ini didasari oleh beberapa alasan:

1. Model ini memiliki *misclassification rate* yang paling kecil dari antara kedua model lainnya
2. Model ini memiliki nilai cumulative lift yang cukup baik untuk hampir sebagian besar persentase data
3. Model ini memiliki nilai *sensitivity* dan 1-*specificity* yang terbaik dari antara kedua model lainnya.

Meskipun Model Forest sebenarnya juga bisa menjadi pertimbangan karena berhasil memprediksi tujuan peminjaman car dengan sangat baik, tetapi belum tentu model ini dapat memprediksi ketujuh tujuan peminjaman yang lain sebaik memprediksi car.

ACKNOWLEDGMENT

Peneliti ingin mengucapkan terima kasih kepada Bapak Iwan Prasetiawan selaku dosen pengampu Mata Kuliah Big Data Analytics. Atas jasa pengajaran dan bimbingan beluai, peneliti dapat menyelesaikan karya ilmiah ini dengan lancar. Peneliti juga ingin mengucapkan terima kasih pula kepada teman-teman yang senantiasa memberi semangat kepada peneliti, sehingga peneliti dapat menyelesaikan karya ilmiah ini dengan tepat waktu

REFERENCES

[1] Y. Religia, A. Nugroho and W. Hadikristanto, "AnalisisPerbandingan Algoritma Optimasi pada Random Forestuntuk KlasifikasiData Bank Marketing," *JURNAL RESTI*, vol. 5, no. 1, p. 189, 2021.

[2] E. Suryati, S. and A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *JURNAL TEKNOLOGI DAN SISTEM INFORMASI*, vol. 4, no. 1, p. 98, 2023.

[3] R. A. Rizal, I. S. Girsang and S. A. Prasetiyo, "KlasifikasiWajah Menggunakan Support Vector Machine (SVM)," *Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 3, no. 2, p. 2, 2019.

[4] D. Darwis, E. S. Pratiwi and F. O. Pasaribu3, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *Jurnal Ilmiah Edutic*, vol. 7, no. 1, p. 2, 2020.

[5] C. Cahyaningtyas, Y. Nataliani and I. R. Widiasari, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," *JURNAL TEKNOLOGI INFORMASI*, vol. 18, no. 2, p. 178, 2021.

[6] A. I. Shafarindu, E. Patimah, Y. M. Siahaan, A. W. Wardhana, B. V. Haekal and D. S. Prasvita, "Klasifikasi Data Penjualan pada Supermarket dengan Metode Decision Tree," *SENAMIKA*, vol. 2, no. 1, p. 661, 2021.