

Nama : Febianus Felix W
NIM : 00000072737
Matakuliah : (IS429-DL) Big Data Analytics - LAB
Pilihan Industri : Banking
Company Name : Bank Uhuy
Ujian Akhir Semester LAB

Repository :

saya menggunakan dataset dari kaggle dengan sumber : <https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset> yang berbentuk CSV, dengan kolom jumlah 17 dan 11,162 baris. dataset mencakup :

Column	Deskripsi Data	
	Data Type	Penjelasan
Age	Double	Umur Pelanggan
Balance	double	Saldo Rekening
Campaign	double	Jumlah kontak yang dilakukan selama kampanye ini dan untuk klien ini.
Contact	string	Komunikasi kontak ('cellular', 'telephone')
Day	double	Hari terakhir kontak (1,2,3....31)
Default	boolean	Apakah memiliki kredit bermasalah? ('no', 'yes', 'unknown')
Deposit	boolean	Apakah klien telah berlangganan deposito berjangka? ('ya', 'tidak')
Duration	double	Durasi kampanye
Education	string	Pendidikan Terakhir Pelanggan
housing	boolean	Apakah memiliki pinjaman perumahan? ('no', 'yes', 'unknown')
job	string	Jenis Pekerjaan Pelanggan ('admin.', 'bluecollar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')

loan	boolean	Apakah memiliki pinjaman pribadi? ('no', 'yes', 'unknown')
marital	string	Status Pernikahan ('divorced', 'married', 'single', 'unknown')
month	string	Bulan terakhir kontak dalam setahun ('jan', 'feb', 'mar', ..., 'nov', 'dec')
pdays	double	Jumlah hari yang berlalu setelah klien terakhir dihubungi dari kampanye sebelumnya.
poutcome	string	Hasil dari kampanye pemasaran sebelumnya ('failure', 'nonexistent', 'success')
previous	double	Jumlah kontak yang dilakukan sebelum kampanye ini dan untuk klien ini.

I. Question 1: CLO021 Sub-CLO-6 and CLO013 Sub-CLO-7 Weight (40 %)

- Manage your **Data Preparation** and **Exploration Process** corresponding to the data curation process.
- The journal writing in section III contains an explanation of EDA (Exploratory Data Analysis)
- **Create your EDA** and **Data preparation** process with the SAS® Data Preparation Application.

Answer :

Pemrosesan data ini dilakukan menggunakan *SAS Visual Analytics*, dimana analisis dilakukan secara eksklusif melalui aplikasi ini. Dengan pendekatan ini, diharapkan dapat meningkatkan efisiensi dalam analisis data secara keseluruhan.

A. EDA

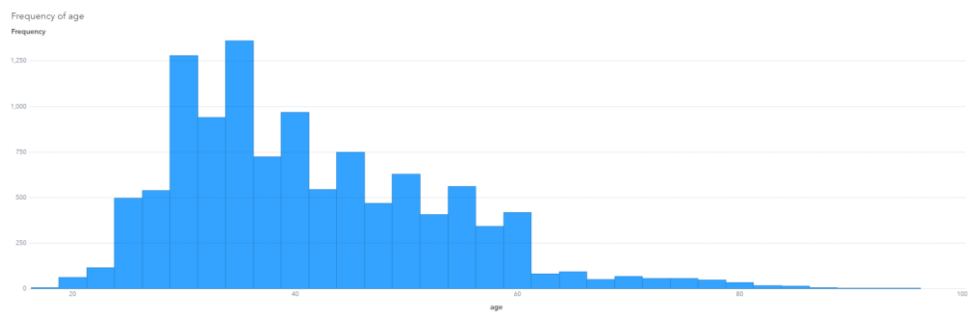
Dengan fitur "Explore and Visualize" dalam SAS Visual Analytics, Bank Uhuy dapat dengan mudah menganalisis data secara mendalam dan menghasilkan visualisasi yang jelas dan informatif. Ini memungkinkan Bank Uhuy untuk melihat tren, pola, dan hubungan dalam data mereka dengan cepat.

pada tahap EDA ini menggunakan Table BANK :

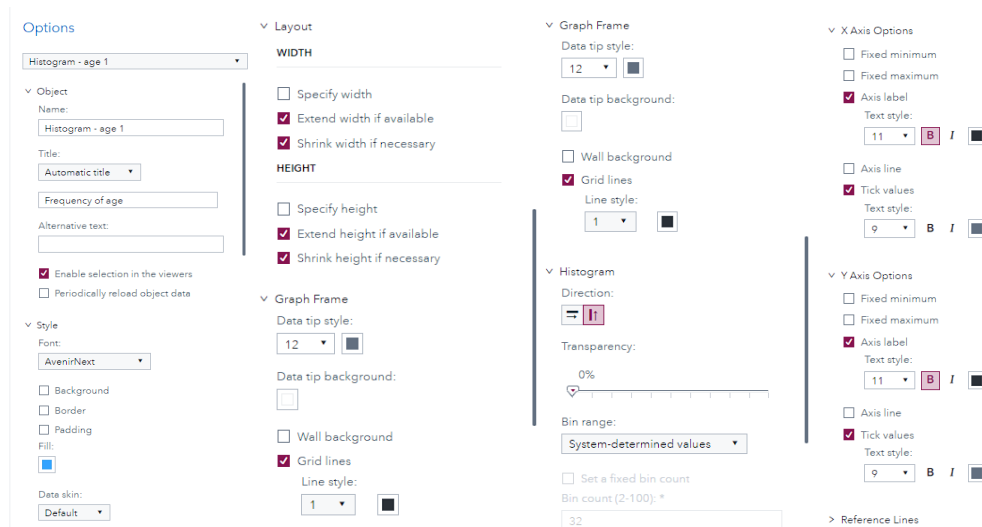


Gambar 1. Dataset Table Bank

1. Bar Plot Variabel Age

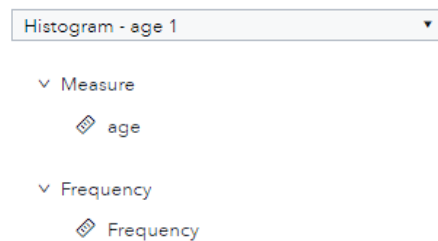


Gambar 2. Bar Chart Variabel Age



Gambar 3. Option Bar Chart Variabel Age

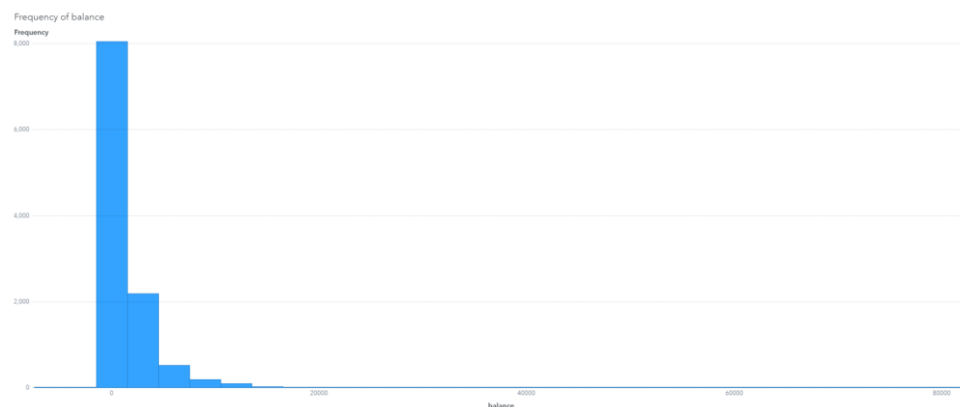
Data Roles



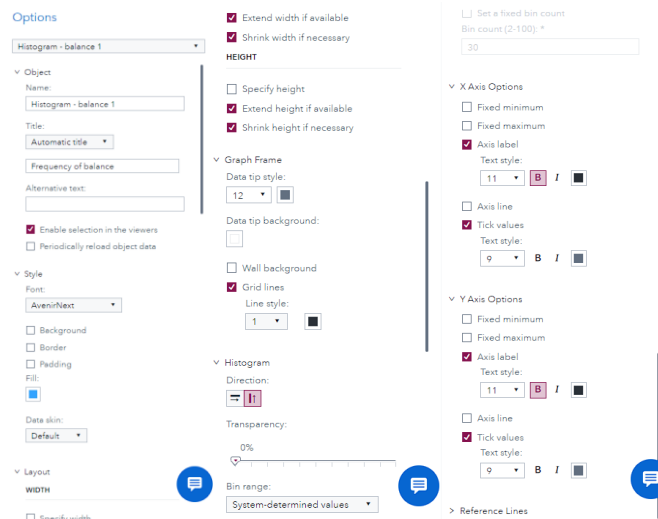
Gambar 4. Roles Bar Chart Variabel Age

Visualisasi ini menggambarkan sebaran variabel age, di mana terlihat bahwa bar plot ini memiliki distribusi Skew-Right. Distribusi ini menunjukkan bahwa mayoritas usia berada di bawah 40 tahun, namun ada juga yang berusia di atas 40 tahun. Pada bar plot ini, mean, median, dan modus juga terlihat. Median data berada pada angka 39, dengan nilai mean sebesar 41.23, dan modus sebesar 31. Dari gambar ini dapat diambil insight bahwa Bank Uhu dapat memberikan direct marketing yang lebih sesuai dengan umur karena mayoritas masih berada dalam usia produktif.

2. Bar Plot Variabel Balance

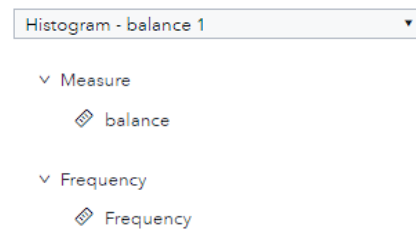


Gambar 5. Bar Chart Variabel Balance



Gambar 6. Option Bar Chart Variabel *Balance*

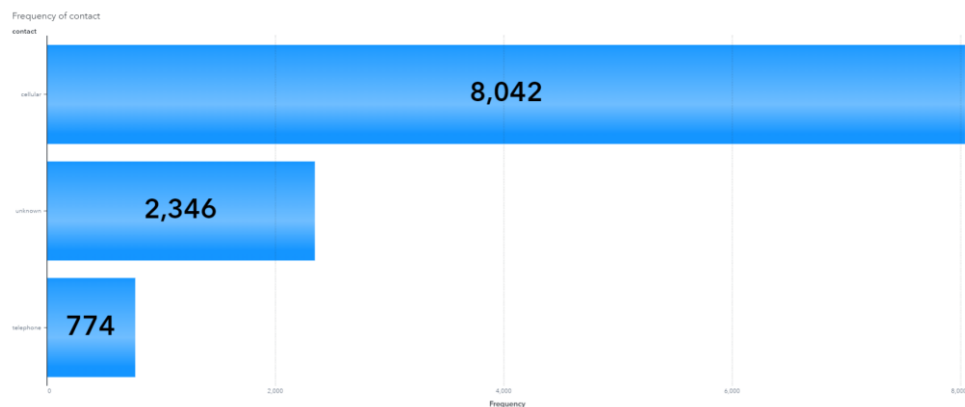
Data Roles



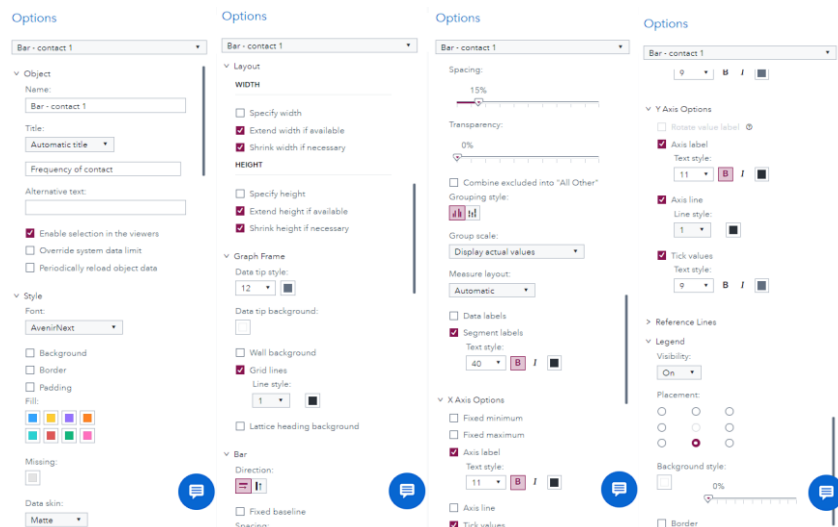
Gambar 7. Roles Bar Chart Variabel *Balance*

Visualisasi ini menggambarkan sebaran variabel Balance, di mana terlihat bahwa mayoritas balance berada di angka 0. Hal ini sejalan dengan tujuan utamanya, yaitu mempromosikan (direct marketing) kepada pelanggan yang belum memiliki keanggotaan bank atau memiliki keanggotaan bank tetapi tidak menaruh uangnya di bank. Median data tersebut terlihat berada pada angka 550, dengan nilai mean sebesar 1,528.54. Dari gambar ini dapat diambil insight bahwa Bank Uhuy dapat meningkatkan strategi pemasaran langsungnya untuk menarik lebih banyak pelanggan agar menyimpan uang mereka di bank.

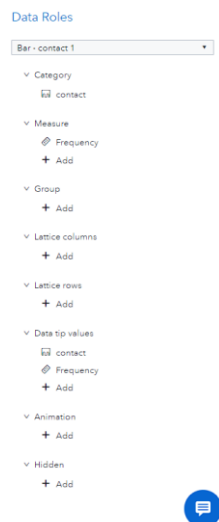
3. Bar Plot Variabel Contact



Gambar 8. Bar Chart Variabel *Contact*



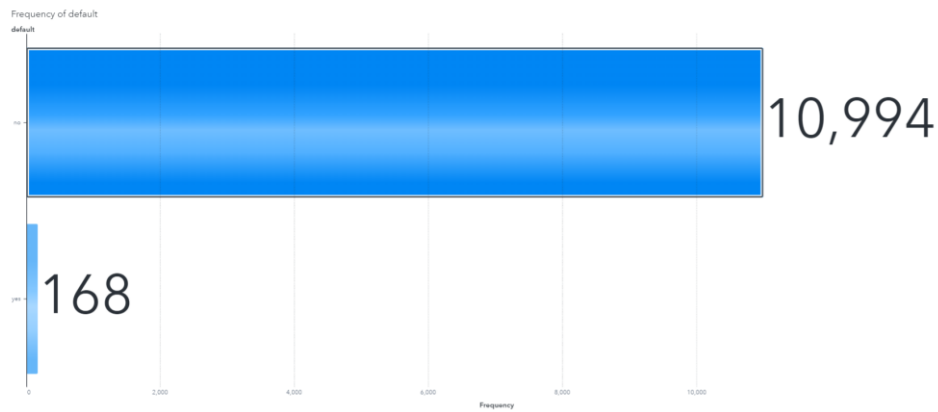
Gambar 9. Option Bar Chart Variabel *Contact*



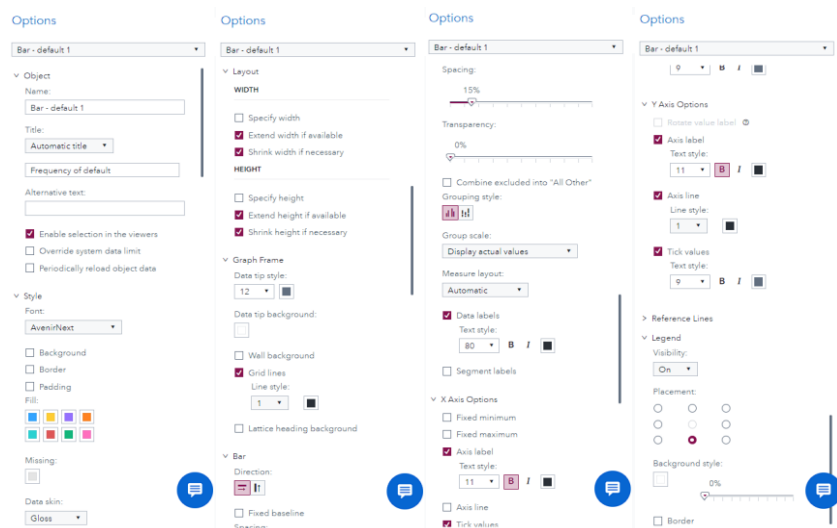
Gambar 10. Roles Bar Chart Variabel *Contact*

Visualisasi ini menggambarkan sebaran variabel *Contact*, di mana terlihat bahwa pelanggan yang dihubungi menggunakan alat komunikasi yang berbeda, yaitu "Cellular", "Unknown", dan "Telephone", dengan jumlah masing-masing 8.042 untuk "Cellular", 2.346 untuk "Unknown", dan 774 untuk "Telephone". Nilai "Unknown" dalam variabel ini dapat mewakili berbagai jenis kontak, seperti melalui email atau kontak langsung. Insight dari visualisasi ini bagi Bank Uhu adalah bahwa sebagian besar pelanggan lebih sering dihubungi melalui telepon seluler, sehingga strategi pemasaran yang berfokus pada komunikasi melalui seluler mungkin lebih efektif.

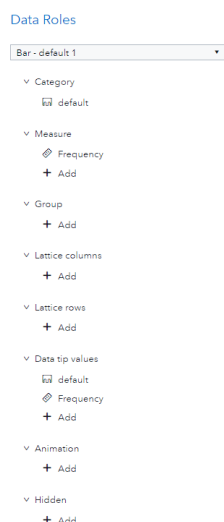
4. Bar Plot Variabel Default



Gambar 11. Bar Chart Variabel *Default*



Gambar 12. Option Bar Chart Variabel *Default*

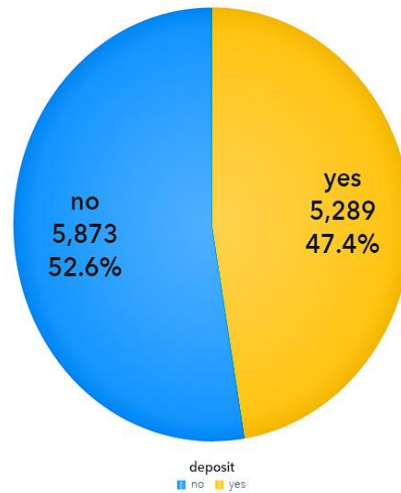


Gambar 13. Roles Bar Chart Variabel *Default*

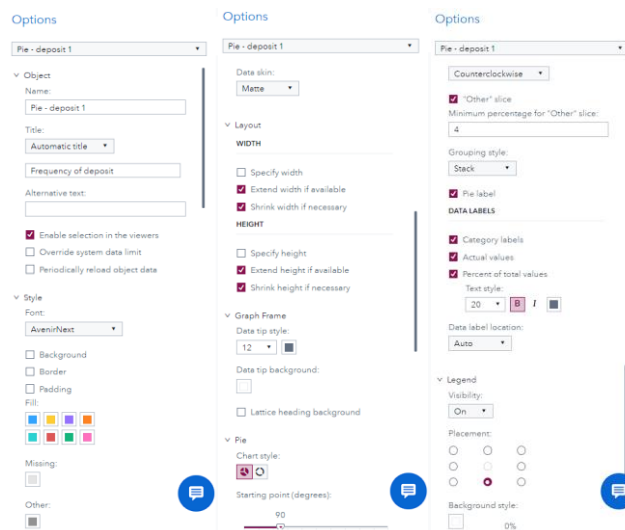
Visualisasi ini menggambarkan sebaran variabel Default, yang memiliki dua nilai berbeda, yaitu "yes" dan "no", dengan jumlah masing-masing 10.994 untuk "no" dan 168 untuk "yes". Variasi nilai "Default" menunjukkan bahwa mayoritas pelanggan tidak memiliki catatan tunggakan, namun ada jumlah yang signifikan yang memiliki

catatan tunggakan. Insight dari visualisasi ini bagi Bank Uhuy adalah bahwa meskipun sebagian besar pelanggan tidak memiliki catatan tunggakan, perlu adanya perhatian khusus terhadap pelanggan dengan catatan tunggakan.

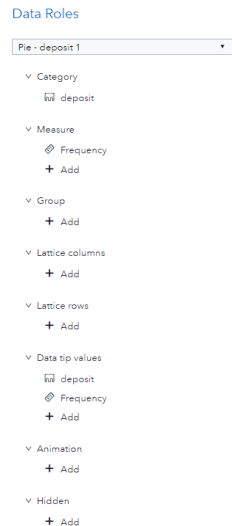
5. Pie Chart Variabel Deposit



Gambar 14. Pie Chart Variabel *Deposit*



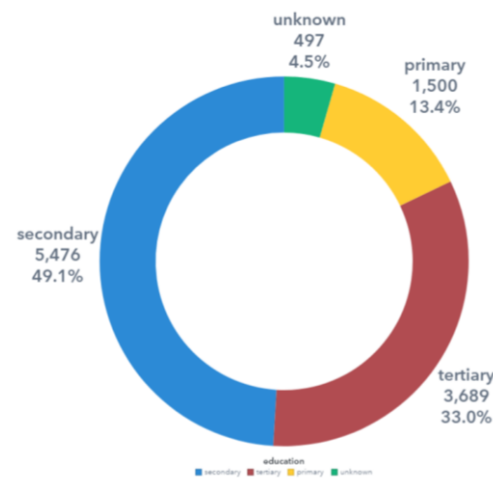
Gambar 15. Option Pie Chart Variabel *Deposit*



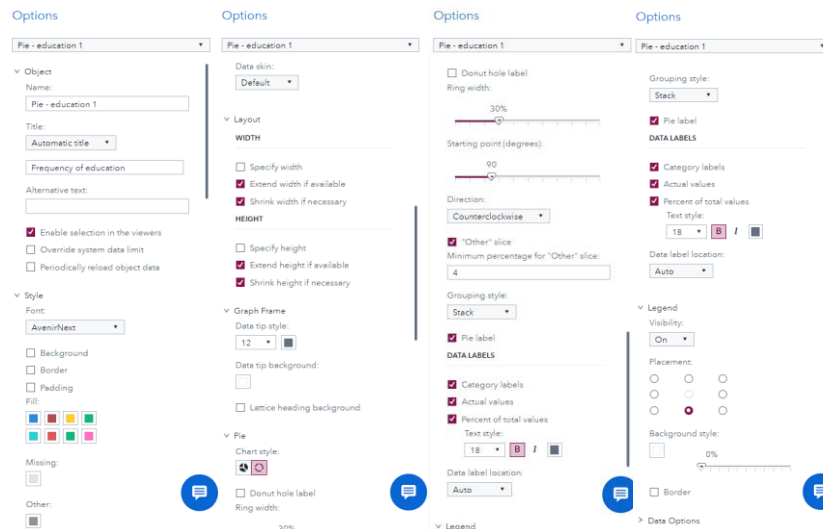
Gambar 16. Roles Pie Chart Variabel *Deposit*

Visualisasi ini menggambarkan pembagian variabel deposit, yang menjadi target dari penelitian ini. Variabel ini menunjukkan apakah setelah direct marketing, pelanggan akan melakukan deposit ke bank atau tidak. Variabel ini terbagi menjadi dua nilai, yaitu "yes" dan "no", dengan jumlah masing-masing sebesar 5.873 (52,6%) untuk "yes" dan 5.289 (47,4%) untuk "no". Variasi nilai "Deposit" menunjukkan tingkat penetrasi produk deposit yang baik di Bank Uhuy. Insight untuk Bank Uhuy adalah bahwa strategi direct marketing yang dilakukan sudah cukup efektif, dengan lebih dari setengah pelanggan memilih untuk melakukan deposit.

6. Pie Chart Variabel Education

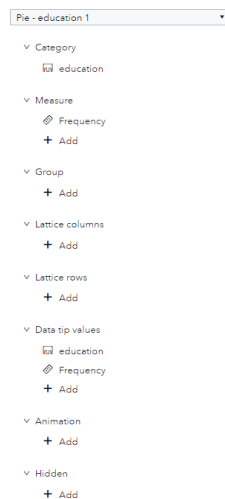


Gambar 17. Pie Chart Variabel *Education*



Gambar 18. Option Pie Chart Variabel *Education*

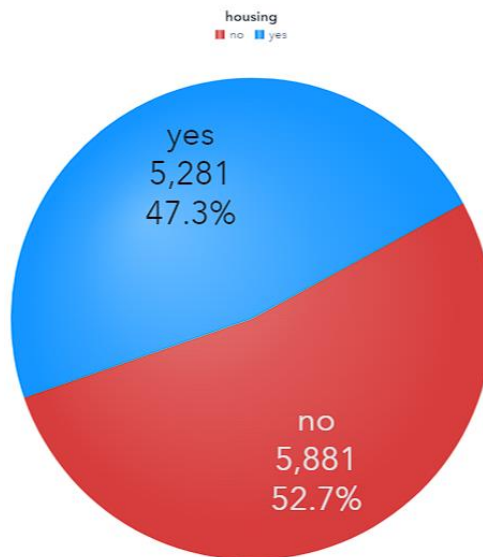
Data Roles



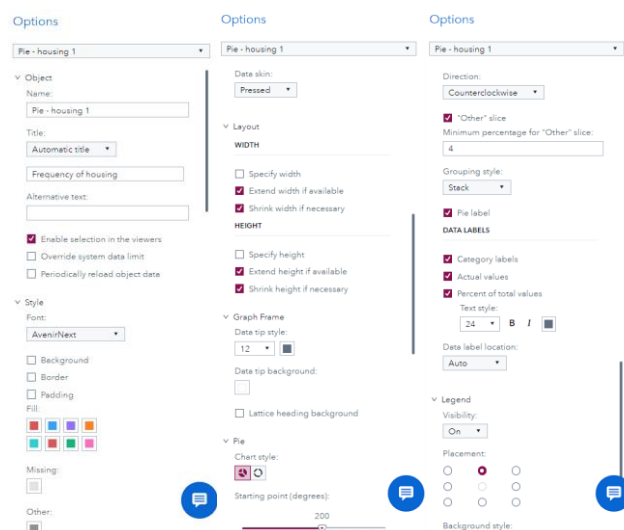
Gambar 19 Roles Pie Chart Variabel *Education*

Visualisasi ini menggambarkan pembagian variabel *Education*, dimana terlihat bahwa variabel ini terbagi tiga, yaitu "Primary", "Secondary", "Tertiary", dan "Unknown", dengan jumlah masing-masing sebesar 1,500 (13.4%) untuk "Primary", 5,476 (49.1%) untuk "Secondary", 3,689 (33%) untuk "Tertiary", dan 497 (4.5%) untuk "Unknown". Insight untuk Bank Uhu dari visualisasi ini adalah bahwa mayoritas pelanggan memiliki tingkat pendidikan "Secondary" dan "Tertiary". Ini menunjukkan bahwa strategi pemasaran dapat disesuaikan untuk lebih mengakomodasi kebutuhan dan preferensi kelompok pelanggan ini. Bank Uhu dapat mempertimbangkan untuk mengembangkan produk dan layanan yang lebih canggih dan sesuai dengan tingkat pendidikan yang lebih tinggi, serta kampanye pemasaran yang lebih mendetail dan informatif.

7. Pie Chart Variabel Housing

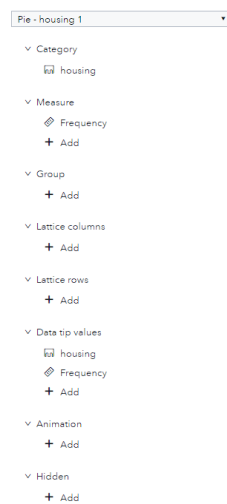


Gambar 20. Pie Chart Variabel *Housing*



Gambar 21. Option Pie Chart Variabel *Housing*

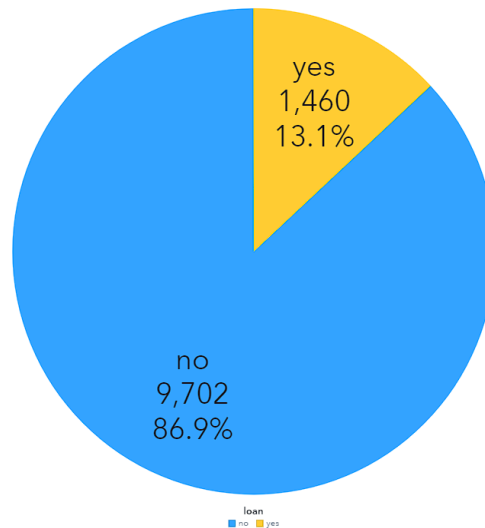
Data Roles



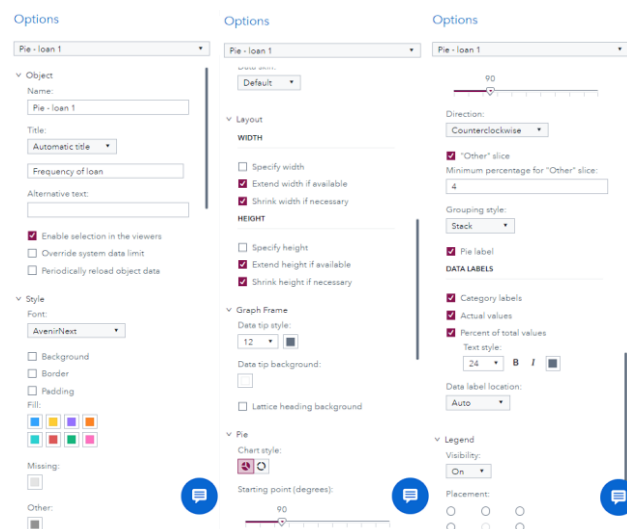
Gambar 22. Roles Pie Chart Variabel *Housing*

Visualisasi ini menggambarkan pembagian variabel Housing. variabel "housing" menunjukkan dua nilai yang berbeda, yaitu "yes" dan "no", dengan jumlah masing-masing sebesar 5,281 (47.3%) untuk "yes", dan 5,881 (52.7%) untuk "no". Variasi nilai dalam variabel "housing" menunjukkan bahwa sebagian besar pelanggan atau prospek Bank Uhuy tidak memiliki pinjaman rumah. insight bagi bank uhuy adalah bahwa ada peluang signifikan untuk menawarkan produk pinjaman rumah kepada lebih dari setengah pelanggan atau prospek yang belum memiliki pinjaman rumah.

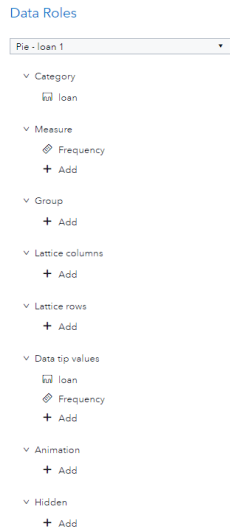
8. Pie Chart Variabel Loan



Gambar 23. Pie Chart Variabel *Loan*



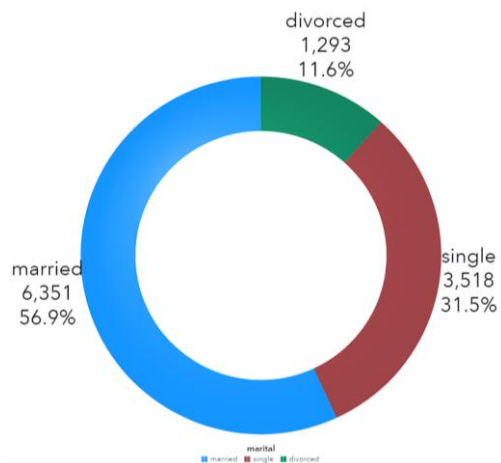
Gambar 24. Option Pie Chart Variabel *Loan*



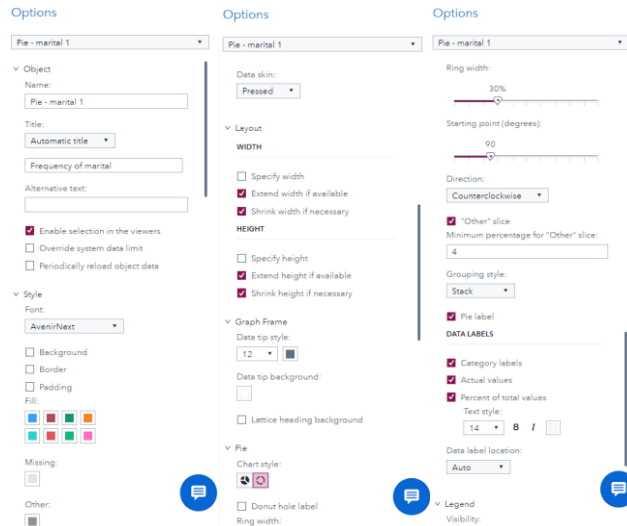
Gambar 25. Roles Pie Chart Variabel *Loan*

Visualisasi ini menggambarkan pembagian variabel Loan. terlihat bahwa variabel "Loan" menunjukkan dua nilai yang berbeda, yaitu "yes" dan "no", dengan jumlah masing-masing sebesar 1,460 (13.1%) untuk "yes", dan 9,702 (86.9%) untuk "no". Variasi nilai dalam variabel "Loan" menunjukkan bahwa sebagian besar pelanggan atau prospek Bank Uhuy tidak memiliki pinjaman. Insight bagi Bank Uhuy adalah bahwa ada peluang besar untuk menawarkan produk pinjaman kepada sebagian besar pelanggan atau prospek yang belum memiliki pinjaman. Bank Uhuy dapat mengembangkan strategi pemasaran yang menargetkan kelompok ini dengan menawarkan berbagai jenis pinjaman yang sesuai dengan kebutuhan dan preferensi pelanggan, serta dengan menawarkan suku bunga yang kompetitif dan fleksibel.

9. Pie Chart Variabel Marital

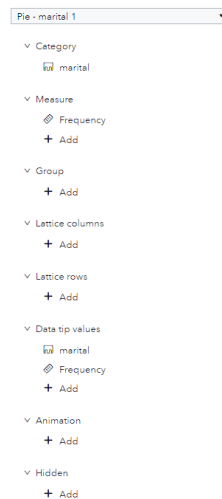


Gambar 26. Pie Chart Variabel *Marital*



Gambar 27. Option Pie Chart Variabel *Marital*

Data Roles



Gambar 28. Roles Pie Chart Variabel *Marital*

Visualisasi ini menggambarkan pembagian variabel Marital. dapat diamati bahwa variabel "Marital" menampilkan empat nilai yang berbeda, yakni "Married", "Single", dan "Divorced", dengan jumlah masing-masing sebesar 6,351 (56.9%) untuk "Married", 3,518 (31.5%) untuk "Single", dan 1,293 (11.6%) untuk "Divorced". Variasi nilai dalam variabel "Marital" menunjukkan distribusi status pernikahan pelanggan atau prospek di Bank Uhu, dengan mayoritas merupakan pasangan yang menikah. Insight bagi Bank Uhu adalah bahwa profil demografis pelanggan atau prospek mereka didominasi oleh pasangan yang menikah. Hal ini dapat menjadi informasi yang berharga dalam mengembangkan strategi pemasaran yang ditargetkan, karena kebutuhan dan preferensi finansial dapat berbeda antara kelompok-kelompok ini.

B. Data Preparation

Data preparation dilakukan untuk Menyiapkan dan mengolah data yang akan digunakan dalam analisis kampanye pemasaran Bank Uhu.

1. Check Missing Values

```

Code
1  /* Memuat data */
2  data bank;
3      infile '/shared/home/febianus.felix@student.umn.ac.id/casuser/bank.csv' delimiter=',';
4  input age job $ marital $ education $ default $ balance housing $ loan $ contact $ day
5  run;
6
7  /* Menghitung jumlah nilai yang hilang dan total pengamatan */
8  proc means data=bank n nmiss;
9      output out=missing_summary(drop=_TYPE_ _FREQ_) nmiss= / autoname;
10 run;
11
12 /* Menghitung total pengamatan */
13 proc sql noprint;
14     select count(*) into :Total_Observations from bank;
15 quit;
16
17 /* Menghitung persentase nilai yang hilang */
18 data missing_percentage;
19     set missing_summary;
20     array nums _numeric_;
21     do over nums;
22         nums = (nums / &Total_Observations) * 100;
23     end;
24 run;
25
26 /* Menampilkan hasil persentase nilai yang hilang */
27 proc print data=missing_percentage;
28 run;
29

```

Gambar 29. Code Check Missing Values

results

Output Data

The MEANS Procedure

Variable	N	N Miss
age	11162	0
balance	11162	0
day	11162	0
duration	11162	0
campaign	11162	0
pdays	11162	0
previous	11162	0

Obs	age_NMiss	balance_NMiss	day_NMiss	duration_NMiss	campaign_NMiss	pdays_NMiss	previous_NMiss
1	0	0	0	0	0	0	0

Gambar 30. Hasil Check Missing Values

Pada langkah ini, pengecekan Missing Values dilakukan terhadap seluruh variabel dalam dataset ini. Pengecekan ini bertujuan untuk memastikan kualitas dan kelengkapan data yang akan digunakan dalam analisis. Dengan identifikasi dan penanganan data yang hilang, diharapkan bias atau kesalahan dalam hasil analisis dapat diminimalisir. Pada gambar di atas, terlihat bahwa tidak ada missing values pada dataset ini, sehingga data siap digunakan untuk analisis lebih lanjut tanpa memerlukan penanganan tambahan terhadap data yang hilang.

2. Handling Outlier

Pada langkah ini, pengecekan outlier dilakukan dengan tujuan untuk diidentifikasi dan ditangani data yang memiliki nilai ekstrim. Proses ini penting karena outlier dapat mempengaruhi hasil analisis secara signifikan dan menyebabkan bias dalam model prediktif.

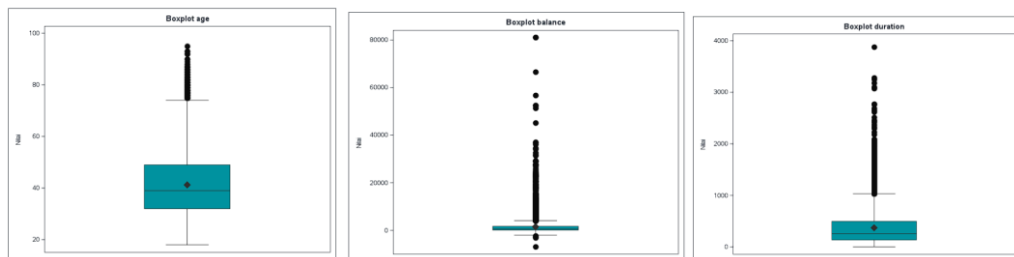
```

/* Membuat makro untuk membuat boxplot */
%macro create_boxplots(data=, columns=);
    ods graphics on;
    %let n = %sysfunc(countw(&columns));
    %do i = 1 %to &n;
        %let col = %scan(&columns, &i);
        proc sgplot data=&data;
            vbox &col;
            title "Boxplot &col";
            yaxis label="Nilai";
            run;
        %end;
    ods graphics off;
%mend create_boxplots;

/* Membuat boxplot setelah penanganan outlier */
%create_boxplots(data=bank, columns=&outlier_columns);

```

Gambar 31. Code Check Outlier



Gambar 32. Hasil Sebelum Remove Outlier

Pada gambar diatas, tampak bahwa terdapat outlier yang signifikan pada variabel age, balance, dan duration. Penanganan outlier dilakukan menggunakan metode IQR, dengan hasil akhir berupa pemindahan outlier ke nilai mean. Langkah ini dilakukan untuk mengurangi dampak negatif dari nilai ekstrim tersebut, sehingga analisis dapat menghasilkan model yang lebih akurat dan bebas bias.

```

/* Membuat makro untuk mengganti outlier dengan nilai batas (lower/upper bound) */
%macro replace_outliers_with_iqr(data=, columns=, iqr_data=);
    data &data;
    set &data;
    if _N_ = 1 then set &iqr_data;
    array cols[*] &columns;
    array q1s[*] age_Q1 balance_Q1 duration_Q1;
    array q3s[*] age_Q3 balance_Q3 duration_Q3;
    do i = 1 to dim(cols);
        lower_bound = q1s[i] - 1.5*(q3s[i] - q1s[i]);
        upper_bound = q3s[i] + 1.5*(q3s[i] - q1s[i]);
        if cols[i] < lower_bound then cols[i] = lower_bound;
        else if cols[i] > upper_bound then cols[i] = upper_bound;
    end;
    drop i lower_bound upper_bound;
run;
%mend replace_outliers_with_iqr;

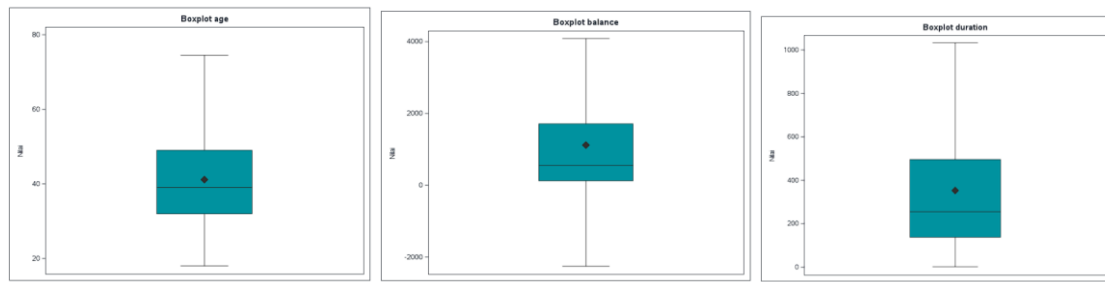
/* Memanggil makro untuk mengganti outlier dengan nilai batas (lower/upper bound) */
%replace_outliers_with_iqr(data=bank, columns=&outlier_columns, iqr_data=iqr_summary);

/* Menampilkan data setelah penggantian outlier */
proc print data=bank(obs=10);
run;

/* Membuat boxplot setelah penanganan outlier */
%create_boxplots(data=bank, columns=&outlier_columns);

```

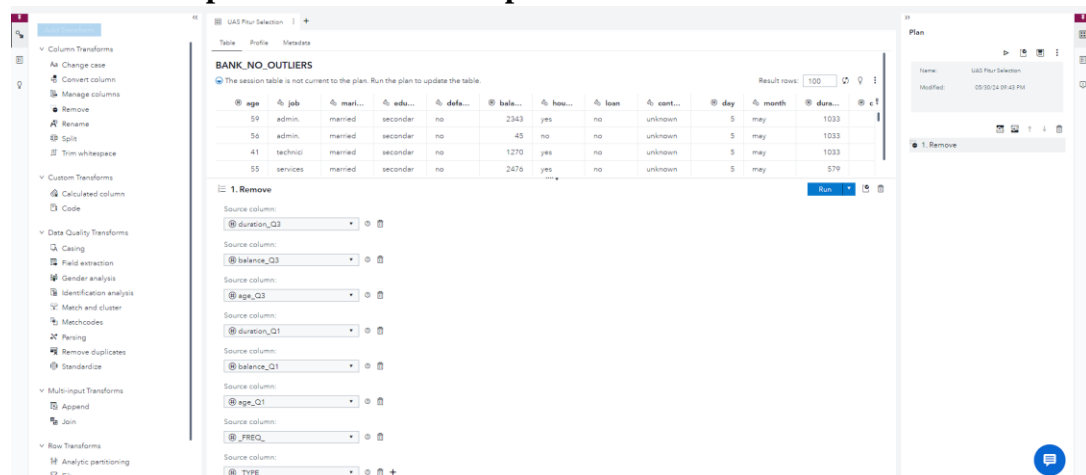
Gambar 33. Code Hapus Outlier



Gambar 34. Hasil Setelah Remove Outlier

Setelah outlier ditangani, dapat terlihat bahwa distribusi data menjadi lebih terkonsentrasi di sekitar nilai mean atau median, dengan rentang nilai yang lebih konsisten dan tidak ada nilai ekstrim yang mencolok.

3. Remove Unimportant & Feature Importance



Gambar 35. Remove variabel yang tidak perlu

Setelah menangani pemeriksaan Missing Values dan Outlier, terdapat kolom tambahan yang dimasukkan ke dalam dataset yang dibutuhkan untuk langkah sebelumnya. Namun, untuk langkah berikutnya dalam analisis, kolom-kolom tersebut tidak lagi diperlukan. Oleh karena itu, diperlukan pemilihan fitur atau variabel yang relevan untuk dipertahankan dalam proses analisis berikutnya. Proses ini bertujuan untuk menyaring variabel yang paling berpengaruh atau relevan terhadap tujuan analisis yang ditetapkan, sehingga dapat meningkatkan efisiensi dan akurasi model yang dibangun.

Penghapusan kolom "Duration_Q3", "Balance_Q3", "Age_Q3", "Duration_Q1", "Balance_Q1", "Age_Q1", "_FREQ_", dan "_Type_" dari dataset menghasilkan pengurangan dimensi dataset, yang pada gilirannya mengurangi kompleksitas analisis data. Dengan mengurangi jumlah variabel yang tidak relevan atau tidak diperlukan, proses analisis menjadi lebih efisien dan fokus pada fitur-fitur yang lebih signifikan untuk mencapai tujuan analisis yang ditetapkan. Hal ini juga dapat meningkatkan kinerja model yang dibangun dengan mengurangi kelebihan variabilitas yang tidak memberikan kontribusi signifikan terhadap hasil analisis.

2. Question 2: CLO021 Sub-CLO-11, CLO021 Sub-CLO-12 and CLO013 Sub-CLO-14 Weight (60 %)

- Build and manage your **Analytics model** using SAS® Software
- The journal writing in section IV contains an explanation of Modeling your data
- **Create your Analytical Model** by design and concept of SAS® VDMML with **Forest or SVM** algorithms **with comparison of other algorithms**.

**** So, in this case you are asked to make at least two algorithms for your model

Jawab :

Data yang saya gunakan dalam modelling adalah data yang telah dibersihkan yaitu Bank_Fix:



Gambar 36. Dataset Table Bank

A. Modelling

1. Train Validation Test Split

×

Edit Partition

Name:

Training Validation Test Split

Based on:

☐ Data item

☒ Sampling

Sampling method:

Simple random sampling ▾

Number of partitions:

3 ▾

Training partition sampling percentage: *

70

Testing partition sampling percentage: *

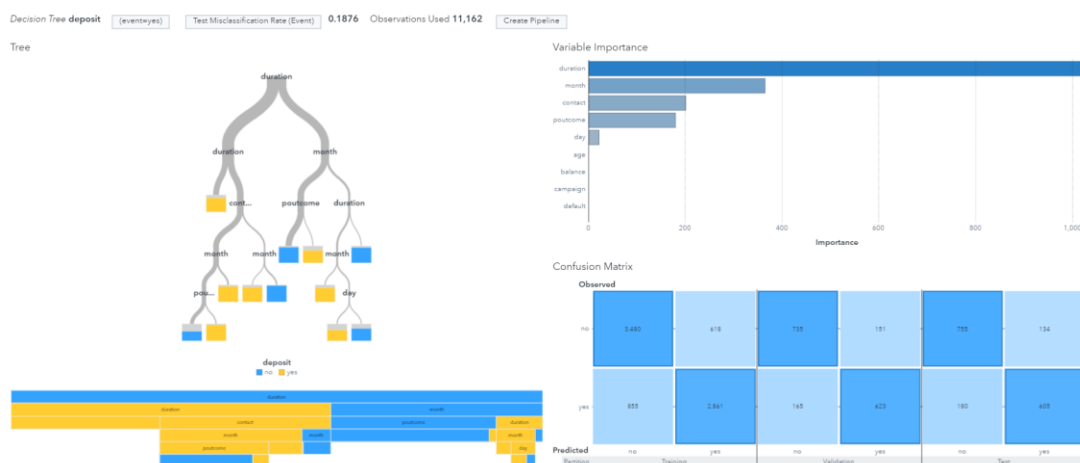
15

☐ Random number seed

OK

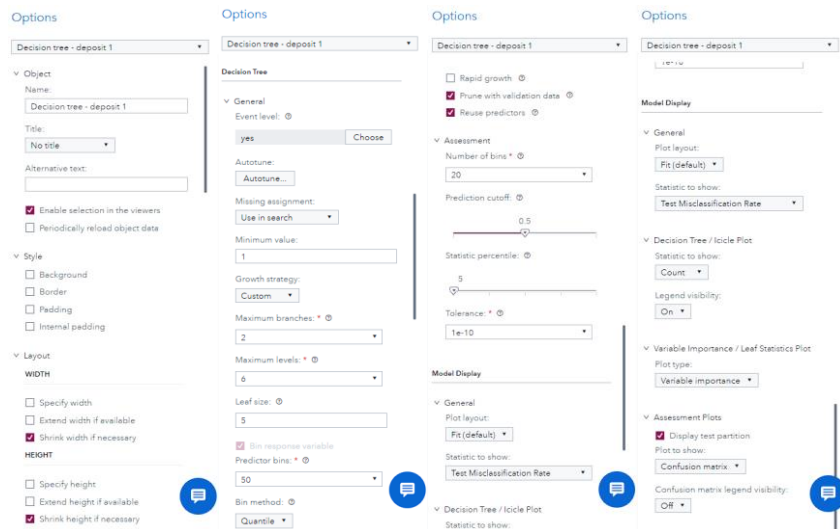
Cancel

Dalam gambar terlihat bahwa pemisahan dilakukan ke dalam tiga bagian, dengan proporsi 70% untuk data training, 15% untuk data testing, dan 15% untuk data validation. Hal ini membantu dalam penyesuaian parameter oleh mesin dan pembelajaran pola-pola dalam data training, diikuti oleh pengujian kinerja model pada data testing, serta validasi hasilnya pada data validation.



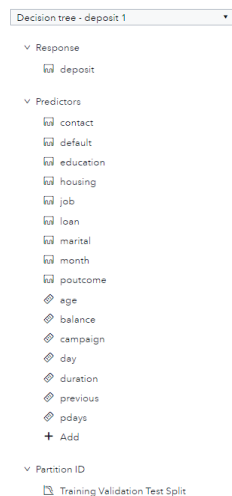
Partition	Misclassification Rate	F1 Score
Training	0.1911	0.789
Validation	0.1864	0.796
Test	0.1906	0.790

Gambar 39. Hasil Pemodelan Decision Tree



Gambar 40. Option Decision Tree

Data Roles



Gambar 41. Roles Decision Tree

Pada pengujian dengan model Decision Tree, yang menghasilkan Misclassification Rate sebesar 0.1965 dan F1-Score sebesar 0.785 pada data test, terdapat hasil yang konsisten dengan confusion matrix. Misclassification Rate yang rendah, ditunjukkan oleh jumlah TP dan TN yang tinggi, menandakan bahwa model jarang melakukan kesalahan dalam prediksi, seperti yang terlihat pada data training dengan TP sebesar 3541 dan TN sebesar 2807. Selain itu, F1-Score yang tinggi menunjukkan keseimbangan yang baik antara precision dan recall, sesuai dengan data validation yang menunjukkan TP sebesar 766 dan TN sebesar 576. Dengan hasil tersebut, model Decision Tree mempertahankan performa yang konsisten dan mampu mengklasifikasikan data dengan baik, sebagaimana yang tercermin dalam hasil pengujian tes dengan TP sebesar 759 dan TN sebesar 582.

3. Random Forest



Gambar 42. Visualisasi Random Forest

Partition	Misclassification Rate	F1 Score
Training	0.1917	0.782
Validation	0.1864	0.789
Test	0.1876	0.786

Gambar 43. Hasil Pemodelan Random Forest

Options

Forest - deposit 1

Object: Name: Forest - deposit 1 Title: No title Alternative text:

☒ Enable selection in the viewers ☐ Periodically reload object data

Style

☐ Background ☐ Border ☐ Padding ☐ Internal padding

Layout

WIDTH

☐ Specify width ☐ Extend width if available ☒ Shrink width if necessary

HEIGHT

☐ Specify height ☐ Extend height if available ☒ Shrink height if necessary

Forest

General

Event level: ☐ yes ☐ Choose

Autotune: Autotune...

Number of trees: 100

Bootstrap: ☐ ☐ 0.6

Vote: ☐ Majority

Splitting criterion: ☐ Information gain ratio

☐ Set fixed number of predictors to split nodes

Number of predictors to split nodes: 4

Decision Tree

Missing assignment: Use in search

Minimum value: 1

Maximum branches: 2

Maximum levels: 6

Leaf size: 5

Predictor bins: 50

Bin method: Quantile

Assessment

Number of bins: 20

Prediction cutoff: 0.5

Statistic percentile: 5

Tolerance: 1e-10

Model Display

General

Plot layout: Fit (default)

Statistic to show: Test Misclassification Rate

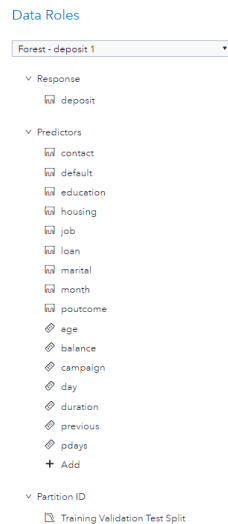
Assessment Plots

☒ Display test partition

Plot to show: Confusion matrix

Confusion matrix legend visibility: On

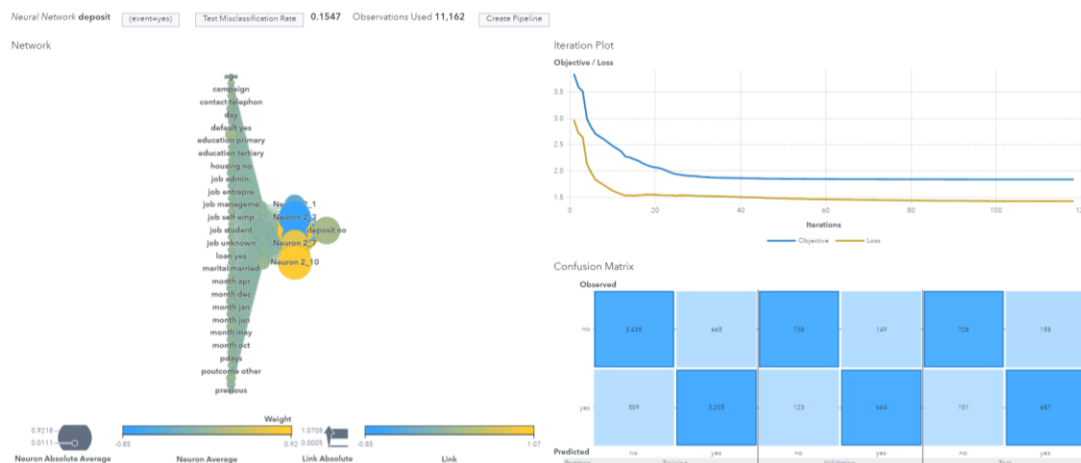
Gambar 44. Option Pemodelan Random Forest



Gambar 45. Roles Pemodelan Random Forest

Pada pengujian dengan model Random Forest, yang menghasilkan Misclassification Rate sebesar 0.1959 dan F1-Score sebesar 0.780 pada data test, terdapat korelasi yang konsisten dengan hasil confusion matrix. Misclassification Rate yang rendah, yang tercermin dari jumlah TP dan TN yang tinggi, menunjukkan bahwa model jarang melakukan kesalahan dalam prediksi, seperti yang terlihat pada data training dengan TP sebesar 3694 dan TN sebesar 2616. Selain itu, F1-Score yang tinggi menunjukkan keseimbangan yang baik antara precision dan recall, sebagaimana yang tercermin dalam data validation yang menunjukkan TP sebesar 792 dan TN sebesar 533. Dengan hasil tersebut, model Random Forest mempertahankan performa yang konsisten dan mampu mengklasifikasikan data dengan baik, seperti yang diamati dalam hasil pengujian tes dengan TP sebesar 788 dan TN sebesar 544.

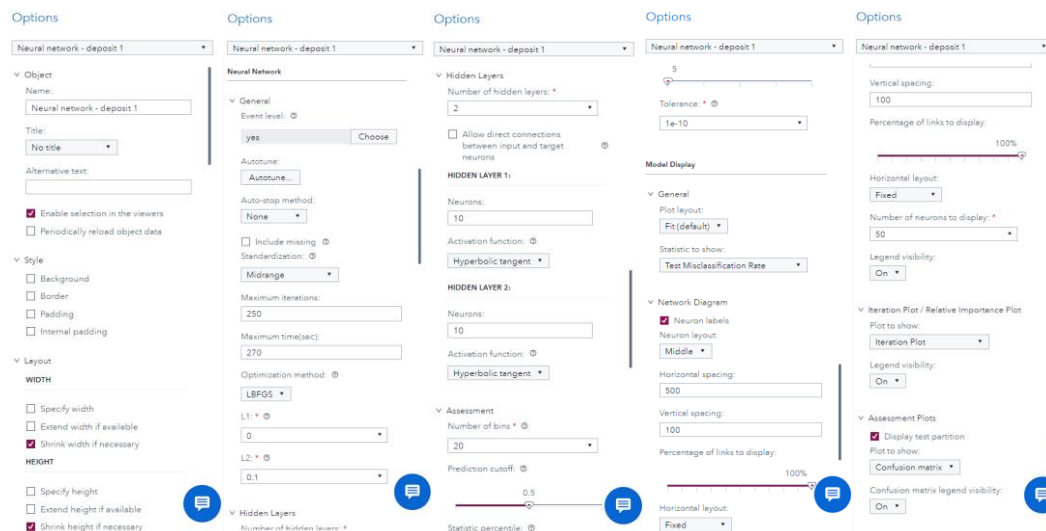
4. Neural Network



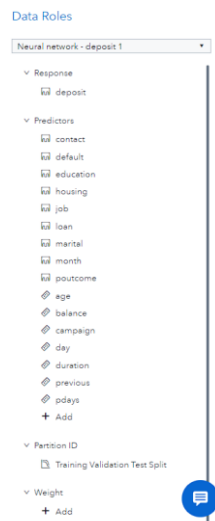
Gambar 46. Visualisasi Neural Network

Partition	Misclassification Rate	F1 Score
Training	0.1515	0.844
Validation	0.1422	0.855
Test	0.1547	0.840

Gambar 47. Visualisasi Neural Network



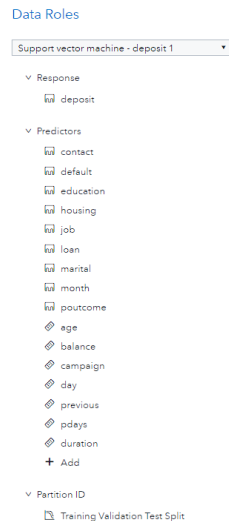
Gambar 48. Visualisasi Neural Network



Gambar 49. Visualisasi Neural Network

Pada pengujian dengan model Neural Network, diperoleh Misclassification Rate sebesar 0.1750 dan F1-Score sebesar 0.813 pada data test. Hasil ini konsisten dengan analisis dari confusion matrix, di mana Misclassification Rate yang rendah tercermin dari jumlah TP dan TN yang tinggi, seperti pada data training dengan TP sebesar 3435 dan TN sebesar 3205. F1-Score yang tinggi juga menunjukkan keseimbangan yang baik antara precision dan recall, sebagaimana yang diamati pada data validation dengan TP sebesar 738 dan TN sebesar 664. Dengan demikian, model Neural Network menunjukkan performa yang baik dan konsisten, sebagaimana terlihat dalam hasil pengujian tes dengan TP sebesar 728 dan TN sebesar 687.

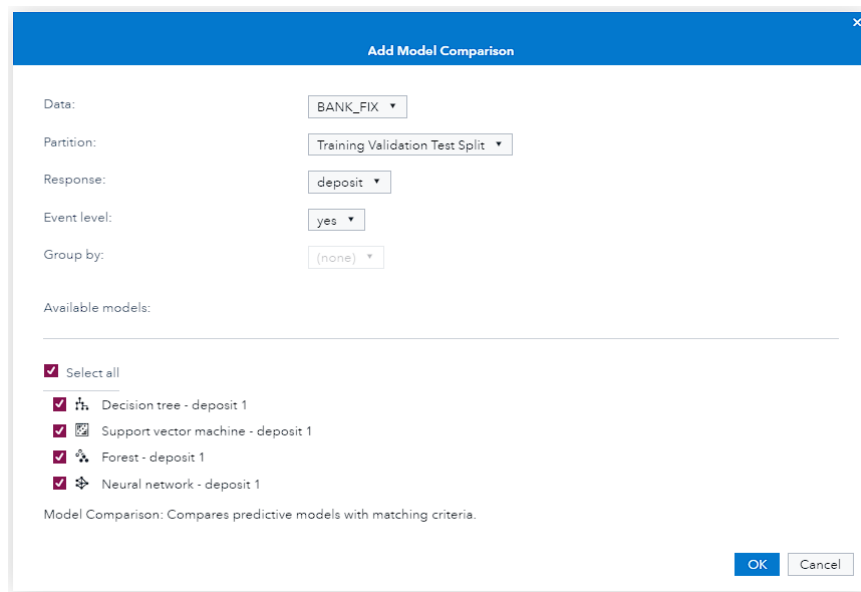
5. Support Vector Machine



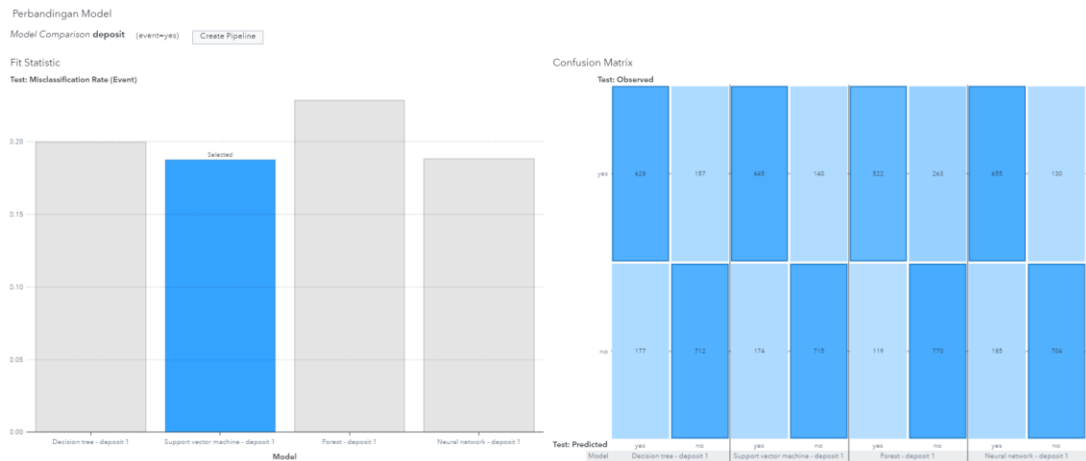
Gambar 53. Option Pemodelan SVM

Pada pengujian dengan model Support Vector Machine (SVM), didapatkan Misclassification Rate sebesar 0.1798 dan F1-Score sebesar 0.806 pada data test. Hasil tersebut mencerminkan performa yang baik dari model, yang diperkuat oleh analisis confusion matrix. Misclassification Rate yang rendah, seperti terlihat pada data training dengan jumlah FP dan FN yang relatif kecil (masing-masing 667 dan 628), menunjukkan bahwa model jarang membuat kesalahan dalam prediksi. Hal ini juga tercermin dalam F1-Score yang tinggi, mengindikasikan keseimbangan yang baik antara precision dan recall. Pada data validation, ditemukan Misclassification Rate sebesar 0.1786 dan F1-Score sebesar 0.810, menunjukkan konsistensi performa model selama proses validasi. Hasil pengujian tes juga menegaskan kualitas model SVM, dengan TP sebesar 655 dan TN sebesar 734. Dengan demikian, model SVM menunjukkan kinerja yang baik dan konsisten dalam mengklasifikasikan data.

B. Model Comparison



Gambar 54. Pemilihan *Model Comparison*



Gambar 55. Visualisasi *Model Comparison*

Selected	Model	Visualization Type	Misclassification Rate (Event)	F1 Score
Yes	Training: Support vector machine - deposit 1	Support Vector Ma...	0.171	0.818
Yes	Validation: Support vector machine - deposit 1	Support Vector Ma...	0.163	0.828
Yes	Test: Support vector machine - deposit 1	Support Vector Ma...	0.161	0.830
No	Training: Decision tree - deposit 1	Decision Tree	0.193	0.790
No	Validation: Decision tree - deposit 1	Decision Tree	0.177	0.807
No	Test: Decision tree - deposit 1	Decision Tree	0.186	0.796
No	Training: Forest - deposit 1	Forest	0.185	0.796
No	Validation: Forest - deposit 1	Forest	0.172	0.811
No	Test: Forest - deposit 1	Forest	0.189	0.790
No	Training: Neural network - deposit 1	Neural Network	0.169	0.823
No	Validation: Neural network - deposit 1	Neural Network	0.170	0.824
No	Test: Neural network - deposit 1	Neural Network	0.166	0.826

Gambar 56. Hasil *Model Comparison*

Options

Model comparison 1

Object

Name: Model comparison 1

Title: Custom title

Perbandingan Model

Alternative text:

☒ Enable selection in the viewers

☐ Periodically reload object data

Style

☐ Background

☐ Border

☐ Padding

☐ Internal padding

Layout

WIDTH

☐ Specify width

☐ Extend width if available

☒ Shrink width if necessary

HEIGHT

☐ Specify height

☐ Extend height if available

Options

Model comparison 1

☒ Shrink height if necessary

Model Comparison

Prediction cutoff: 0.5

Percentile: 5

Comparison Display

Plot layout: Fit (default)

Fit Statistics

Fit statistic: Test F1 Score

Assessment Plots

☐ Show all plots

Plot to show: Confusion matrix

Partition: Training

Confusion matrix legend visibility: On

Gambar 57. Option *Model Comparison*

Dalam pengujian yang melibatkan empat model utama, yaitu Decision Tree, Random Forest, Neural Network, dan Support Vector Machine (SVM), dilakukan analisis menyeluruh terhadap kinerja masing-masing model. Misclassification Rate sebesar 0.1965 dan F1-Score 0.785 dihasilkan oleh Decision Tree, menunjukkan performa yang konsisten dengan hasil confusion matrix yang menunjukkan jaranganya kesalahan prediksi. Random Forest juga menunjukkan performa yang baik dengan Misclassification Rate 0.1959 dan F1-Score 0.780, serta memiliki keseimbangan yang

baik antara precision dan recall. Namun, Neural Network menonjol sebagai model terbaik dengan Misclassification Rate terendah sebesar 0.1750 dan F1-Score tertinggi sebesar 0.813. Hasil ini konsisten dengan analisis confusion matrix yang menunjukkan performa konsisten dan tingginya akurasi prediksi. Model SVM juga menunjukkan performa yang baik dengan Misclassification Rate sebesar 0.1798 dan F1-Score sebesar 0.806, serta konsistensi dalam hasil confusion matrix.

Secara keseluruhan, Neural Network adalah model yang paling unggul dalam mengklasifikasikan data, diikuti oleh SVM, Decision Tree, dan Random Forest secara berurutan. Evaluasi ini memberikan panduan yang solid dalam pemilihan model yang dapat diandalkan untuk aplikasi praktis, terutama yang membutuhkan tingkat akurasi tinggi dalam pengklasifikasian data.

URL SHARE

<https://v4e080.vfe.sas.com/links/resources/report?uri=%2Freports%2Freports%2Fdf95e87-1a41-4af1-bc55-49f45198a208&page=vi3764>