

Laporan Kerja Individu

Final Project - Data Analytics

Program Zenius Studi Independen Bersertifikat - Angkatan 4

Nama Lengkap: Dinny Meilinda Sari

Nomor ID live class: 082

Nomor Kelompok: 30

Mentor: Erwin Fernanda

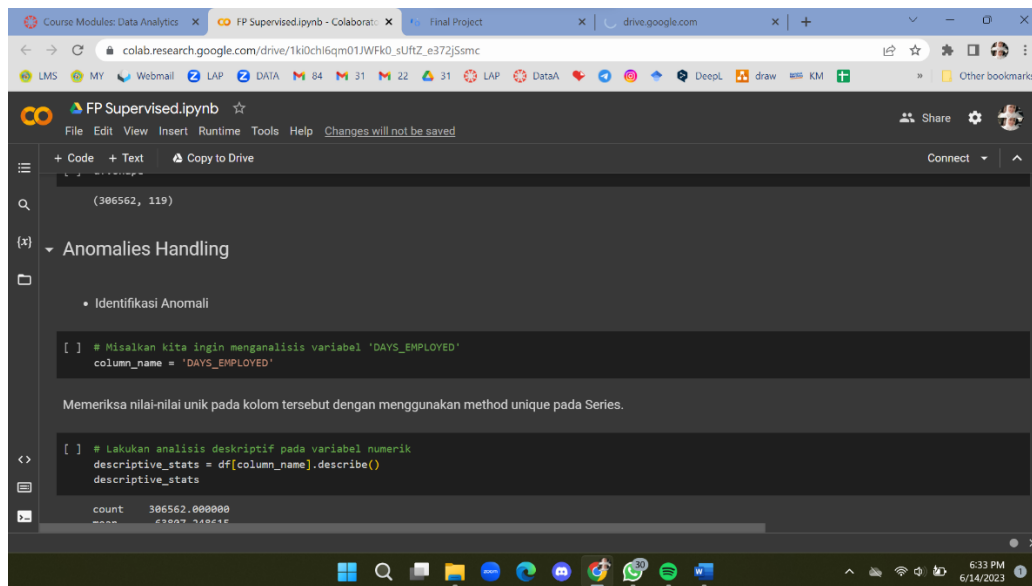
Deskripsi peran

Deskripsikan peran/pekerjaan kamu dalam pengerjaan final project secara singkat namun jelas.

Tulis dalam poin-poin (1 poin 1 kalimat).

- Mengerjakan *Anomalies* dan *Outlier*
- Mengerjakan modeling *Supervised Learning* menggunakan *Decision Tree*
- Membuat dashboard di Looker Studio

Lampiran Hasil Kerja



The screenshot shows a Jupyter Notebook interface with the following content:

```
(386562, 119)
```

Anomalies Handling

- Identifikasi Anomali

```
[ ] # Misalkan kita ingin menganalisis variabel 'DAYS_EMPLOYED'
column_name = 'DAYS_EMPLOYED'
```

Memeriksa nilai-nilai unik pada kolom tersebut dengan menggunakan method unique pada Series.

```
[ ] # Lakukan analisis deskriptif pada variabel numerik
descriptive_stats = df[column_name].describe()
descriptive_stats
```

count 386562.000000
max 386562.000000

```
Course Modules: Data Analytics x FP Supervised.ipynb - Colaboratory x Final Project - Slide 1 x drive.google.com x +
colab.research.google.com/drive/1ki0chl6qm01jWfK0_sUftZ_e372j5smc

LMS MY Webmail LAP DATA B4 31 22 31 LAP DataA DeepL draw KM Other bookmarks

FP Supervised.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive Connect ^

[] # Misalkan kita ingin mengatasi anomali pada variabel 'DAYS_EMPLOYED'
column_name = "DAYS_EMPLOYED"

Pertama, nilai yang dianggap anomali (dalam hal ini nilai yang bernilai 365243) diganti dengan nilai modus dengan menggunakan loc.

[] df.loc[df[column_name] == 365243, column_name] = -200

• Verifikasi Hasil

[] # Periksa nilai-nilai unik pada variabel
pd.set_option('display_max_rows', None) # Display all rows
value_counts = df[column_name].value_counts()
print(value_counts)

-200    55353
-224     312
-210     151
-199     150
-212     149
-220     142
-184     142
-211     140
-202     138
-216     137
-186     136
-188     136
-116     136
```

```
Course Modules: Data Analytics x FP Supervised.ipynb - Colaboratory x Final Project - Slide 1 x drive.google.com x +
colab.research.google.com/drive/1ki0chl6qm01jWfK0_sUftZ_e372j5smc

LMS MY Webmail LAP DATA B4 31 22 31 LAP DataA DeepL draw KM Other bookmarks

FP Supervised.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive Connect ^

array([-637, -1188, -225, ..., -12971, -11084, -8694])

[] # lakukan analisis deskriptif setelah penanganan
descriptive_stats_after = df[column_name].describe()
descriptive_stats_after

count    386562.000000
mean     -1991.373755
std       2278.033413
min      -17912.000000
25%      -2761.000000
50%      -1214.000000
75%       -289.000000
max         9.000000
Name: DAYS_EMPLOYED, dtype: float64

[] # Menghitung persentase perubahan setelah penanganan
percentage_change = (descriptive_stats_after['mean'] - descriptive_stats['mean']) / descriptive_stats['mean'] * 100
print("Percentage Change:", percentage_change)

Percentage Change: 183.12892889528838

• Penanganan Anomali pada Numerik Diskrit

[] # Mencari nilai unik pada kolom 'CNT_CHILDREN'
unique_values = df['CNT_CHILDREN'].unique()

f.1 # Mengganti nilai anomali dengan nilai yang mundkkti
```

```
Course Modules: Data Analytics x FP Supervised.ipynb - Colaboratory x Final Project - Slide 1 x drive.google.com x +
colab.research.google.com/drive/1ki0chl6qm01jWfK0_sUftZ_e372j5smc

LMS MY Webmail LAP DATA B4 31 22 31 LAP DataA DeepL draw KM Other bookmarks

FP Supervised.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive Connect ^

4 Random Forest - Random Over Sampling 1.000000 0.509240 0.509240 0.355179 0.912355 0.020444
5 KNN - Random Over Sampling 0.929436 0.560065 0.560065 0.534306 0.615732 0.322756

Decision Tree

[] import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
import matplotlib.pyplot as plt

Periksa struktur dan konten dataset dengan melihat beberapa baris pertama:

[] print(df.head())

SK_ID_CURR TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OML_CAR \
0 1000002 1 0 1 0
1 1000003 0 0 0 0
2 1000004 0 1 1 1
3 1000006 0 0 0 0
4 1000007 0 0 1 0

FLAG_OML_REALTY CNT_CHILDREN AMT_INCOME_TOTAL AMT_CREDIT AMT_ANNUITY \
0 1 0 202500.0 480597.5 24290.5
1 0 270000.0 1293502.5 35498.5
```

```
Course Modules: Data Analytics x FP Supervised.ipynb - Colaboratory x Final Project - Slide 1 x drive.google.com x +
colab.research.google.com/drive/1ki0chl6qm01/WfK0_sUfZ_e372j5smc

LMS MY Webmail LAP DATA 84 31 22 31 LAP DataA DeepL draw KM + Other bookmarks

FP Supervised.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive
Connect ^

Pisahkan fitur (features) dari label (target):

[ ] X = df.drop('TARGET', axis=1)
    y = df['TARGET']

Lakukan pemrosesan data tambahan, seperti mengisi nilai yang hilang (missing values) dan mengkodekan variabel kategorikal.

Bagi data menjadi data latih (training data) dan data uji (test data):

[ ] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

Membangun Model Decision Tree
  • Inisialisasi objek model Decision Tree:

[ ] model = DecisionTreeClassifier()

Latih model menggunakan data latih:

[ ] model.fit(X_train, y_train)
    DecisionTreeClassifier
```

```
Course Modules: Data Analytics x FP Supervised.ipynb - Colaboratory x Final Project - Slide 1 x drive.google.com x +
colab.research.google.com/drive/1ki0chl6qm01/WfK0_sUfZ_e372j5smc

LMS MY Webmail LAP DATA 84 31 22 31 LAP DataA DeepL draw KM + Other bookmarks

FP Supervised.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive
Connect ^

Tampilkan confusion matrix:

[ ] cm = confusion_matrix(y_test, y_pred)
    print("Confusion Matrix:")
    print(cm)

Confusion Matrix:
[[5134  506]
 [ 418  801]]

Tampilkan laporan klasifikasi (classification report):

[ ] report = classification_report(y_test, y_pred)
    print("Classification Report:")
    print(report)

Classification Report:
precision    recall  f1-score   support

 0       0.93     0.91     0.92     56394
 1       0.14     0.16     0.15      4919

 accuracy          0.53          0.54          0.53     61313
 macro avg          0.53          0.54          0.53     61313
 weighted avg          0.86          0.85          0.86     61313

Visualisasi Decision Tree
```

```
Course Modules: Data Analytics x FP Supervised.ipynb - Colaboratory x Final Project - Slide 1 x drive.google.com x +
colab.research.google.com/drive/1ki0chl6qm01/WfK0_sUfZ_e372j5smc

LMS MY Webmail LAP DATA 84 31 22 31 LAP DataA DeepL draw KM + Other bookmarks

FP Supervised.ipynb
File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive
Connect ^

# Visualisasi Decision Tree
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())

dot: graph is too large for cairo-renderer bitmaps. Scaling by 0.0457511 to fit

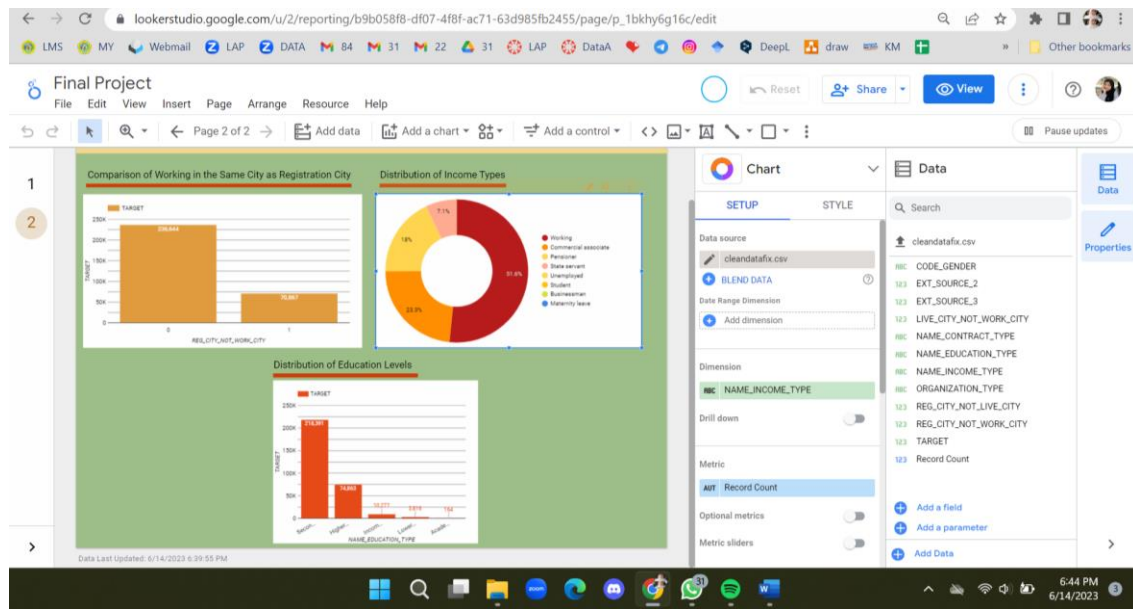
Menggunakan Fitur-Target yang Tergantung (Dependent Features)

  • Dalam beberapa kasus, ada fitur yang secara langsung tergantung pada target atau memiliki korelasi kuat dengan target.
  • Dalam pemodelan Decision Tree, kita dapat menggunakan fitur-fitur ini sebagai pemisah pertama untuk meningkatkan kemampuan prediksal model.

[ ] # Misalnya, kita memiliki fitur 'EXT_SOURCE_1' yang memiliki korelasi kuat dengan target
    dependent_feature = 'EXT_SOURCE_1'

[ ] # Pisahkan fitur-fitur yang tergantung dari fitur-fitur lainnya
    X_dependent = X[dependent_feature]
    X_independent = X.drop(dependent_feature, axis=1)

[ ] # Bagi data menjadi data latih dan data uji
    X_train_dep, X_test_dep, y_train_dep, y_test_dep = train_test_split(X_dependent, y, test_size=0.2, random_state=42)
```



Proses Kerja

No.	Hari, Tanggal	Pekerjaan	Keterangan
1.	Rabu, 7 Juni 2023	Analisis Anomalies dan Outlier	Melakukan eksplorasi data untuk mengidentifikasi dan menganalisis nilai yang tidak biasa atau berbeda secara signifikan dalam dataset yang diberikan.
2.	Sabtu, 10 Juni 2023	Modeling Decision Tree	Mengumpulkan dan membersihkan data, membagi dataset menjadi data latih dan data uji, serta menggunakan algoritma Decision Tree untuk membuat model yang dapat memprediksi hasil yang diinginkan berdasarkan variabel yang ada.
3.	Minggu, 11 Juni 2023	Evaluasi Kinerja Model	Menghitung akurasi, presisi, dan recall untuk mengevaluasi kinerja model Decision Tree yang telah dibuat.
4.	Senin, 12 Juni 2023	Pembuatan Chart Dashboard	Membuat berbagai jenis grafik, seperti bar chart, line chart, donut chart, dan pie chart, untuk menampilkan visualisasi data yang informatif dan menarik.
5.	Selasa, 13 Juni 2023	Pengembangan Dashboard	Menambahkan judul yang jelas dan keterangan yang membantu dalam menginterpretasi data pada setiap chart yang telah dibuat, serta memastikan

			tampilan dashboard mudah dipahami dan menarik secara visual.
6.	Rabu, 14 Juni 2023	Penyelesaian Final Project	Menyelesaikan tugas-tugas terakhir dalam Final Project, meningkatkan kemampuan, dan mengoptimalkan hasil untuk mencapai kesuksesan dalam penyelesaian tugas ini.
7.	Kamis, 15 Juni 2023	Mengedit Video	Mengedit Video Kelompok dengan durasi di bawah 10 menit.

Persetujuan Anggota

Minta anggota kelompok untuk membaca laporan kamu. Jika mereka setuju dengan isi laporan kamu, minta mereka centang checkbox.

Laporan ini telah disetujui oleh:

- ☒ Christ Jordan Baeha
- ☒ Zulfa Nabilah Nurvitasari
- ☒ Dinny Meilinda Sari
- ☐ Sania Salsabila Agustin
- ☐ Fiqih Imanul Haq