```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import plotly.express as px
import plotly.offline as py
import plotly.graph_objs as go
import plotly.tools as tls
import plotly.figure_factory as ff

from collections import Counter
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import cross_val_score, StratifiedKFold, KFold

from sklearn.model_selection import train_test_split, GridSearchCV

from sklearn.metrics import (accuracy_score,
                             classification_report,
                             recall_score, precision_score, f1_score,
                             confusion_matrix)

from xgboost import XGBClassifier
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.ensemble import RandomForestClassifier

!pip install shap
import shap

%matplotlib inline
```

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

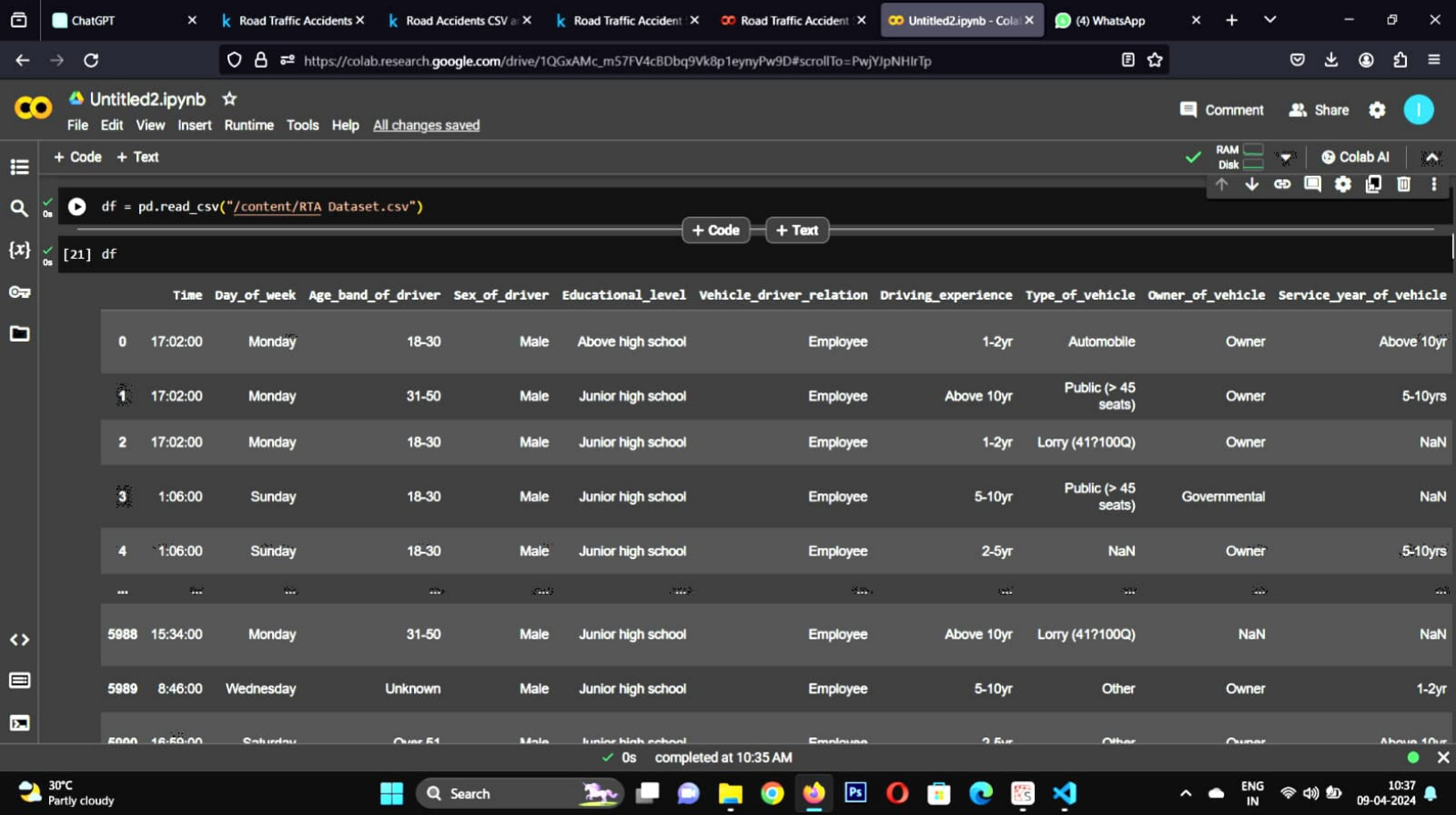
RAM Disk Colab AI

```
!pip install shap
import shap

%matplotlib inline

sns.set_style('darkgrid')
pd. set_option("display.max_columns", None)

import warnings
warnings.filterwarnings('ignore')
```

```
Requirement already satisfied: shap in /usr/local/lib/python3.10/dist-packages (0.45.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from shap) (1.25.2)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from shap) (1.11.4)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from shap) (1.2.2)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from shap) (2.0.3)
Requirement already satisfied: tqdm>=4.27.0 in /usr/local/lib/python3.10/dist-packages (from shap) (4.66.2)
Requirement already satisfied: packaging>20.9 in /usr/local/lib/python3.10/dist-packages (from shap) (24.0)
Requirement already satisfied: slicer==0.0.7 in /usr/local/lib/python3.10/dist-packages (from shap) (0.0.7)
Requirement already satisfied: numba in /usr/local/lib/python3.10/dist-packages (from shap) (0.58.1)
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.10/dist-packages (from shap) (2.2.1)
Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba->shap) (0.41.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->shap) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->shap) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->shap) (2024.1)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->shap) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->shap) (3.4.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas->shap) (1.16.0)
```

```
[20] df = pd.read_csv("/content/RTA Dataset.csv")
```

0s completed at 10:35 AM



ChatGPT | Road Traffic Accidents | Road Accidents CSV | Road Traffic Accident | Road Traffic Accident | Untitled2.ipynb - Cola | (4) WhatsApp

https://colab.research.google.com/drive/1QGxAMc_m57FV4cBDbq9Vk8p1eynyPw9D#scrollTo=PwjYJpNHIrTp

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
df = pd.read_csv("/content/RTA Dataset.csv")
```

+ Code + Text

```
[21] df
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle
0	17:02:00	Monday	18-30	Male	Above high school	Employee	1-2yr	Automobile	Owner	Above 10yr
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	Above 10yr	Public (> 45 seats)	Owner	5-10yrs
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	1-2yr	Lorry (41?100Q)	Owner	NaN
3	1:06:00	Sunday	18-30	Male	Junior high school	Employee	5-10yr	Public (> 45 seats)	Governmental	NaN
4	1:06:00	Sunday	18-30	Male	Junior high school	Employee	2-5yr	NaN	Owner	5-10yrs
...	...	...	...	...	...	...	...	...	...	...
5988	15:34:00	Monday	31-50	Male	Junior high school	Employee	Above 10yr	Lorry (41?100Q)	NaN	NaN
5989	8:46:00	Wednesday	Unknown	Male	Junior high school	Employee	5-10yr	Other	Owner	1-2yr
5990	16:59:00	Saturday	Over 51	Male	Junior high school	Employee	2-5yr	Other	Owner	Above 10yr

0s completed at 10:35 AM

30°C Partly cloudy

Search

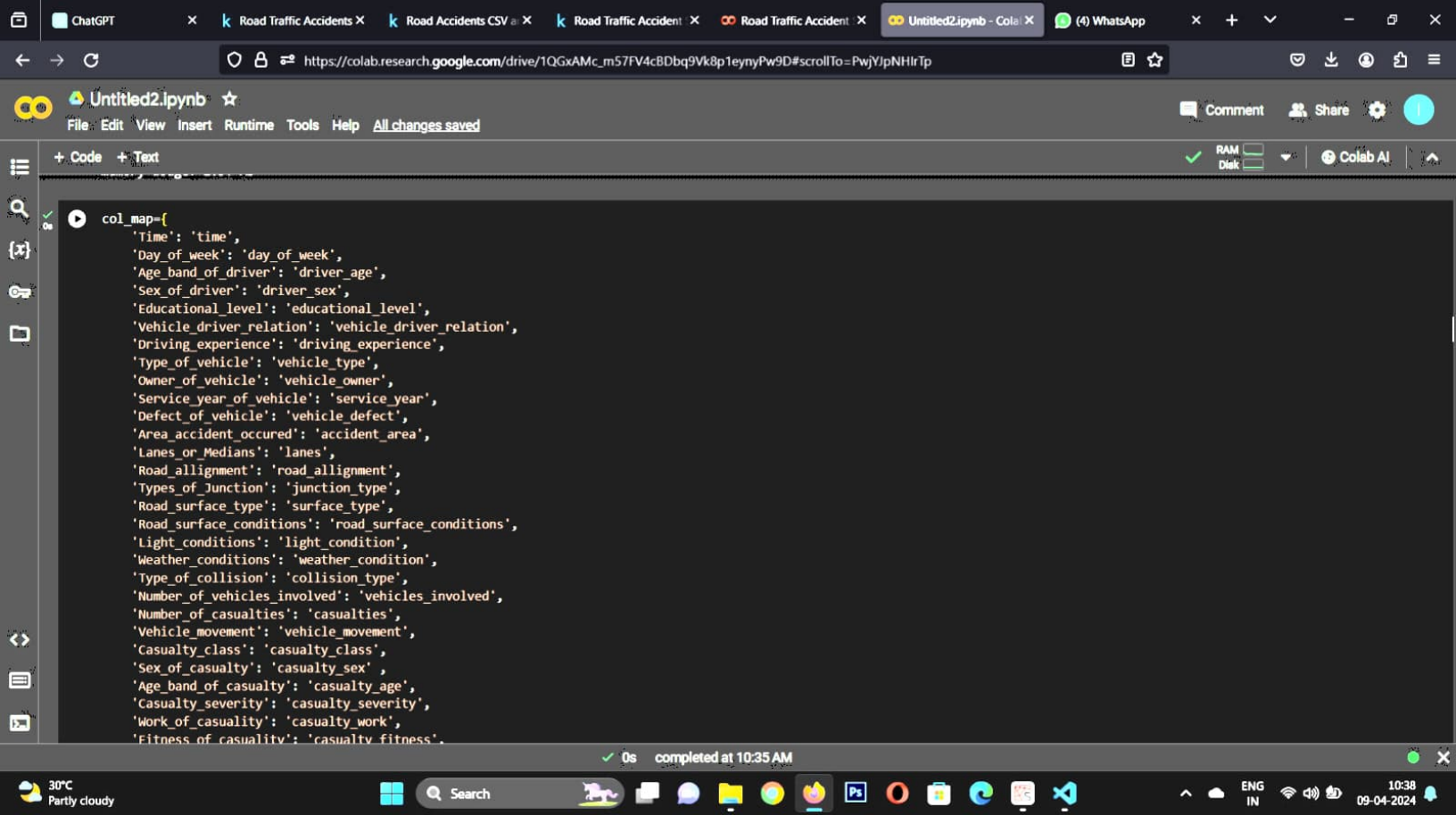
ENG IN 10:37 09-04-2024

```
[22] df.columns
```

```
Index(['Time', 'Day_of_week', 'Age_band_of_driver', 'Sex_of_driver',
       'Educational_level', 'Vehicle_driver_relation', 'Driving_experience',
       'Type_of_vehicle', 'Owner_of_vehicle', 'Service_year_of_vehicle',
       'Defect_of_vehicle', 'Area_accident_occured', 'Lanes_or_Medians',
       'Road_allignment', 'Types_of_Junction', 'Road_surface_type',
       'Road_surface_conditions', 'Light_conditions', 'Weather_conditions',
       'Type_of_collision', 'Number_of_vehicles_involved',
       'Number_of_casualties', 'Vehicle_movement', 'Casualty_class',
       'Sex_of_casualty', 'Age_band_of_casualty', 'Casualty_severity',
       'Work_of_casuality', 'Fitness_of_casuality', 'Pedestrian_movement',
       'Cause_of_accident', 'Accident_severity'],
      dtype='object')
```

```
[23] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5993 entries, 0 to 5992
Data columns (total 32 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   Time                         5993 non-null   object
 1   Day_of_week                  5993 non-null   object
 2   Age_band_of_driver           5993 non-null   object
 3   Sex_of_driver                5992 non-null   object
 4   Educational_level            5643 non-null   object
 5   Vehicle_driver_relation      5746 non-null   object
 6   Driving_experience           5595 non-null   object
 7   Type_of_vehicle              5511 non-null   object
 8   Owner_of_vehicle             5762 non-null   object
```



```
col_map={
    'Time': 'time',
    'Day_of_week': 'day_of_week',
    'Age_band_of_driver': 'driver_age',
    'Sex_of_driver': 'driver_sex',
    'Educational_level': 'educational_level',
    'Vehicle_driver_relation': 'vehicle_driver_relation',
    'Driving_experience': 'driving_experience',
    'Type_of_vehicle': 'vehicle_type',
    'Owner_of_vehicle': 'vehicle_owner',
    'Service_year_of_vehicle': 'service_year',
    'Defect_of_vehicle': 'vehicle_defect',
    'Area_accident_occured': 'accident_area',
    'Lanes_or_Medians': 'lanes',
    'Road_allignment': 'road_allignment',
    'Types_of_Junction': 'junction_type',
    'Road_surface_type': 'surface_type',
    'Road_surface_conditions': 'road_surface_conditions',
    'Light_conditions': 'light_condition',
    'Weather_conditions': 'weather_condition',
    'Type_of_collision': 'collision_type',
    'Number_of_vehicles_involved': 'vehicles_involved',
    'Number_of_casualties': 'casualties',
    'Vehicle_movement': 'vehicle_movement',
    'Casualty_class': 'casualty_class',
    'Sex_of_casualty': 'casualty_sex' ,
    'Age_band_of_casualty': 'casualty_age',
    'Casualty_severity': 'casualty_severity',
    'Work_of_casuality': 'casualty_work',
    'Fitness_of_casuality': 'casualty_fitness',
```

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
df.describe(include=['O']).T
```

	count	unique	top	freq
time	5993	927	16:00:00	57
day_of_week	5993	7	Friday	975
driver_age	5993	6	31-50	2042
driver_sex	5992	3	Male	5483
educational_level	5643	7	Junior high school	3696
vehicle_driver_relation	5746	4	Employee	4630
driving_experience	5595	7	5-10yr	1665
vehicle_type	5511	17	Automobile	1573
vehicle_owner	5762	4	Owner	5088
service_year	4019	6	Unknown	1377
vehicle_defect	3791	3	No defect	3736
accident_area	5874	14	Other	1893
lanes	5793	7	Two-way (divided with broken lines road marking)	2156
road_allignment	5925	9	Tangent road with flat terrain	5109
junction_type	5992	7	Y Shape	2399
surface_type	5911	5	Asphalt roads	5505

✓ 0s completed at 10:35 AM

```
df.describe()
```

	vehicles_involved	casualties
count	5992.000000	5992.000000
mean	1.972964	1.465621
std	0.624651	0.928860
min	1.000000	1.000000
25%	2.000000	1.000000
50%	2.000000	1.000000
75%	2.000000	2.000000
max	6.000000	8.000000

```
[27] df.isna().sum()
```

```
time                      0
day_of_week               0
driver_age                0
driver_sex                1
educational_level       350
vehicle_driver_relation 247
driving_experience      398
vehicle_type            482
vehicle_owner           231
service_year           1974
```

ChatGPT | Road Traffic Accidents | Road Accidents CSV | Road Traffic Accident | Road Traffic Accident | Untitled2.ipynb - Colab | (4) WhatsApp

https://colab.research.google.com/drive/1QGxAMc_m57FV4cBDbq9Vk8p1eynyPw9D#scrollTo=PwjYJpNHIrTp

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[28] # print duplicates
     print("Number of duplicates: ", df.duplicated().sum())
     print("Unique values per column:")
     df.nunique()
```

```
Number of duplicates:  0
Unique values per column:
time                        927
day_of_week                   7
driver_age                    6
driver_sex                    3
educational_level             7
vehicle_driver_relation       4
driving_experience            7
vehicle_type                 17
vehicle_owner                 4
service_year                  6
vehicle_defect                3
accident_area                14
lanes                         7
road_allignment               9
junction_type                 7
surface_type                  5
road_surface_conditions       4
light_condition               4
weather_condition             9
collision_type               10
vehicles_involved             5
casualties                    7
vehicle_movement             13
casualty_class                4
casualty_sex                  3
casualty_age                  6
```

Snipping Tool

Screenshot copied to clipboard and saved
Select here to mark up and share the image

0s completed at 10:35 AM

30°C Partly cloudy

Search

ENG IN 10:39 09-04-2024

ChatGPT | Road Traffic Accidents | Road Accidents CSV | Road Traffic Accident | Road Traffic Accident | Untitled2.ipynb - Colab | (4) WhatsApp

https://colab.research.google.com/drive/1QGxAMc_m57FV4cBDbq9Vk8p1eynyPw9D#scrollTo=PwjYJpNHIrTp

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

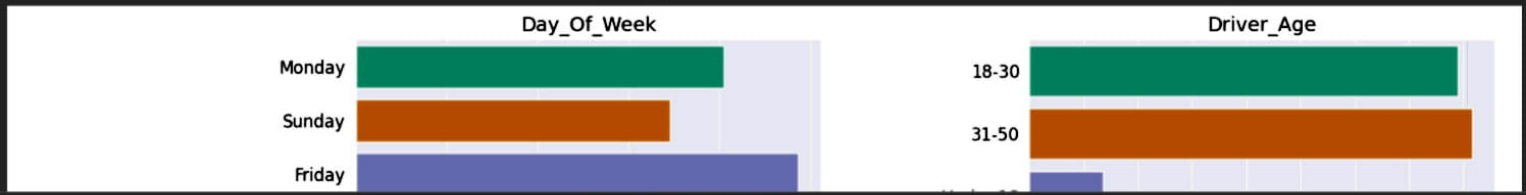
Comment Share

+ Code + Text

RAM Disk Colab AI

[33] 0s
```
# extracting hour and minute from timestamp
df['hour'] = df['time'].dt.hour
df['minute'] = df['time'].dt.minute
df.drop('time', axis=1, inplace=True)
```

14s
```
plt.figure(figsize=(15,70))
plotnumber = 1

for col in df.drop(['hour', 'minute', 'lanes', 'road_allignment', 'pedestrian_movement'], axis=1):
    if plotnumber <= df.shape[1]:
        ax1 = plt.subplot(16,2,plotnumber)
        sns.countplot(data=df, y=col, palette='Dark2')
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
plt.tight_layout()
```



0s completed at 10:35 AM

30°C Partly cloudy

Search

ENG IN 10:39 09-04-2024

Untitled2.ipynb

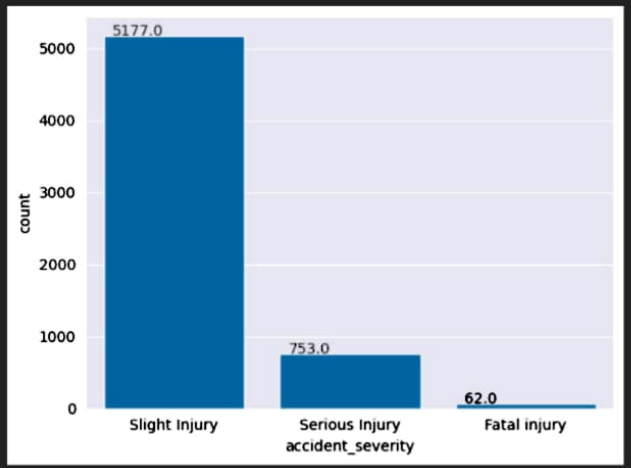
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
ax = sns.countplot(x=df["accident_severity"])
for p in ax.patches:
     ax.annotate(f'{p.get_height()}', (p.get_x()+0.05, p.get_height()+1))

plt.show()
```



0s completed at 10:35 AM

ChatGPT | Road Traffic Accidents | Road Accidents CSV | Road Traffic Accident | Road Traffic Accident | Untitled2.ipynb - Colab | (4) WhatsApp

https://colab.research.google.com/drive/1QGxAMc_m57FV4cBDbq9Vk8p1eynyPw9D#scrollTo=PwjYJpNHIrTp

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[41] def ordinal_encoder(df, feats):
        for feat in feats:
            feat_val = list(1+np.arange(df[feat].nunique()))
            feat_key = list(df[feat].sort_values().unique())
            feat_dict = dict(zip(feat_key, feat_val))
            df[feat] = df[feat].map(feat_dict)
        return df

    df = ordinal_encoder(df, df.drop(['accident_severity'], axis=1).columns)
    df.shape
```

```
(5993, 29)
```

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

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

```
(4195, 28) (1798, 28) (4195,) (1798,)
```

```
[43] counter = Counter(y_train)

    print("============================")

    for k,v in counter.items():
        per = 100*v/len(y_train)
        print(f"Class= {k}, n={v} ({per:.2f}%)")
```

0s completed at 10:35 AM

30°C Partly cloudy

Search

ENG IN 10:40 09-04-2024

```
for k,v in counter.items():
    per = 100*v/len(y_train)
    print(f"Class= {k}, n={v} ({per:.2f}%)")

oversample = SMOTE()
X_train, y_train = oversample.fit_resample(X_train, y_train)

counter = Counter(y_train)

print("============================")

for k,v in counter.items():
    per = 100*v/len(y_train)
    print(f"Class= {k}, n={v} ({per:.2f}%)")

print("============================")

print("Upsampled data shape: ", X_train.shape, y_train.shape)
```

```
============================
Class= Slight Injury, n=3621 (86.32%)
Class= Serious Injury, n=531 (12.66%)
Class= Fatal injury, n=43 (1.03%)
============================
Class= Slight Injury, n=3621 (33.33%)
Class= Serious Injury, n=3621 (33.33%)
Class= Fatal injury, n=3621 (33.33%)
============================
Upsampled data shape:  (10863, 28) (10863,)
```