lr

June 26, 2024

```python
import pandas as pd
import numpy as nm
from sklearn.model_selection import train_test_split as train_test_split
from sklearn.metrics import classification_report
import re
import string
import matplotlib.pyplot as plt
```

```python
data_true=pd.read_csv("/content/drive/MyDrive/train.csv")
data_fake=pd.read_csv("/content/drive/MyDrive/test.csv")
```

```python
from google.colab import drive
drive.mount("/content/drive")
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).

```python
data_true.head(5)
```

```
          ID  Price  Levy Manufacturer    Model  Prod. year   Category  \
0  45654403  13328  1399       LEXUS    RX 450        2010       Jeep
1  44731507  16621  1018    CHEVROLET   Equinox       2011       Jeep
2  45774419   8467     -       HONDA       FIT        2006  Hatchback
3  45769185   3607   862        FORD    Escape        2011       Jeep
4  45809263  11726   446       HONDA       FIT        2014  Hatchback

   Leather interior Fuel type Engine volume    Mileage  Cylinders  \
0              Yes    Hybrid           3.5  186005 km        6.0
1               No    Petrol             3  192000 km        6.0
2               No    Petrol           1.3  200000 km        4.0
3              Yes    Hybrid           2.5  168966 km        4.0
4              Yes    Petrol           1.3   91901 km        4.0

   Gear box type Drive wheels  Doors            Wheel   Color  Airbags
0      Automatic          4x4  04-May      Left wheel  Silver       12
1      Tiptronic          4x4  04-May      Left wheel   Black        8
2       Variator        Front  04-May  Right-hand drive   Black        2
3      Automatic          4x4  04-May      Left wheel   White        0
```

1

```
      4      Automatic          Front  04-May        Left wheel  Silver          4
```

```python
data_fake.head(5)
```

```
         ID Levy Manufacturer   Model  Prod. year   Category Leather interior  \
0  44020629    -   VOLKSWAGEN    Golf        2012  Hatchback               No
1  45784798    -      HYUNDAI  Sonata        2012      Sedan              Yes
2  45800257    -       NISSAN   Tiida        2005      Sedan               No
3  45797981  975        VOLVO    XC90        2012       Jeep              Yes
4  45814303    -         OPEL   Astra        1993  Hatchback               No

  Fuel type Engine volume     Mileage  Cylinders Gear box type Drive wheels  \
0    Diesel     2.0 Turbo        0 km          4        Manual        Front
1    Petrol           2.4    26000 km          4     Tiptronic        Front
2    Petrol           1.5   168000 km          4     Automatic        Front
3    Petrol           3.2   143000 km          6     Automatic          4x4
4    Petrol           1.6   200000 km          4        Manual        Front

    Doors             Wheel      Color  Airbags  Price
0  02-Mar       Left wheel       Grey       10    NaN
1  04-May       Left wheel       Grey       10    NaN
2  04-May  Right-hand drive  Sky blue        8    NaN
3  04-May       Left wheel       Blue       12    NaN
4  04-May       Left wheel      Black        0    NaN
```

```python
data_true.shape,data_fake.shape
```

```
((19237, 18), (8245, 18))
```

```python
data_true["class"]=1
data_fake["class"]=0
```

```python
data_true_manual_testing = data_true.tail(10)
for i in range(385,19,-1):
    data_true.drop([i],axis=0,inplace=True)

data_fake_manual_testing = data_fake.tail(10)
for i in range(385,19,-10):
    data_fake.drop([i],axis=0,inplace=True)
```

```python
data_manual_testing = pd.
    ↪concat([data_fake_manual_testing,data_true_manual_testing])
data_manual_testing.to_csv("manual_testing.csv")
```

```python
data_merge = pd.concat([data_fake,data_true],axis=0)
data_merge.head(10)#displaying first 10
```

```
          ID  Levy    Manufacturer   Model  Prod. year   Category  \
0   44020629     -      VOLKSWAGEN    Golf         2012  Hatchback
1   45784798     -         HYUNDAI  Sonata         2012      Sedan
2   45800257     -          NISSAN   Tiida         2005      Sedan
3   45797981   975           VOLVO    XC90         2012       Jeep
4   45814303     -            OPEL   Astra         1993  Hatchback
5   45812490   639             BMW     320         2014      Sedan
6   45772399  1646           LEXUS  GX 460         2015       Jeep
7   45766222  1172   MERCEDES-BENZ   E 350         2011      Sedan
8   45803362   891         HYUNDAI  Sonata         2016      Sedan
9   45809004   645          DAEWOO Lacetti         2011      Sedan

   Leather interior Fuel type Engine volume      Mileage  Cylinders  \
0                No    Diesel     2.0 Turbo         0 km        4.0
1               Yes    Petrol           2.4     26000 km        4.0
2                No    Petrol           1.5    168000 km        4.0
3               Yes    Petrol           3.2    143000 km        6.0
4                No    Petrol           1.6    200000 km        4.0
5               Yes    Petrol     2.0 Turbo     59733 km        4.0
6               Yes    Petrol           4.6    102907 km        8.0
7               Yes    Diesel           3.5     64496 km        6.0
8               Yes       LPG             2    300234 km        4.0
9               Yes    Petrol           1.8    170000 km        4.0

   Gear box type Drive wheels   Doors            Wheel     Color  Airbags  \
0         Manual        Front  02-Mar       Left wheel      Grey       10
1      Tiptronic        Front  04-May       Left wheel      Grey       10
2      Automatic        Front  04-May  Right-hand drive  Sky blue        8
3      Automatic          4x4  04-May       Left wheel      Blue       12
4         Manual        Front  04-May       Left wheel     Black        0
5      Tiptronic         Rear  04-May       Left wheel     Black       12
6      Automatic          4x4  04-May       Left wheel     Black        0
7      Automatic          4x4  04-May       Left wheel      Blue       12
8      Automatic        Front  04-May       Left wheel    Silver        4
9      Automatic        Front  04-May       Left wheel     White        4

   Price  class
0    NaN      0
1    NaN      0
2    NaN      0
3    NaN      0
4    NaN      0
5    NaN      0
6    NaN      0
7    NaN      0
8    NaN      0
9    NaN      0
```

```python
[75]: print(data_merge.groupby(['Manufacturer'])['Model'].count())
      data_merge.groupby(['Manufacturer'])['Model'].count().plot(kind="bar")
      plt.title("Articles per subject",size=20)
      plt.xlabel("Category",size=20)
      plt.ylabel("Article count",size=15)
      plt.show()
```

```
Manufacturer
ACURA              17
ALFA ROMEO          4
ASTON MARTIN        2
AUDI              246
BENTLEY             1
                  ...
VAZ                51
VOLKSWAGEN        559
VOLVO              17
ZAZ                 3
                    2
Name: Model, Length: 64, dtype: int64
```
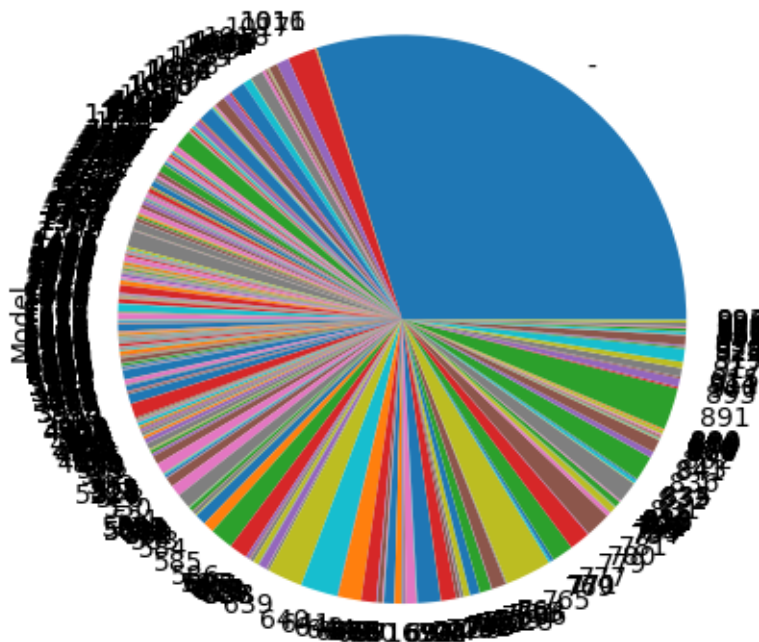
Articles per subject


```
print(data_merge.groupby(['Levy'])['Model'].count())
print("0 = Fake news\n1 = True news")
data_merge.groupby(['Levy'])['Model'].count().plot(kind="pie")
plt.title("Fake news and true news", size=20)
plt.show
```

```
Levy
-        5101
1011       12
1016        4
1017      274
1018      121
          ...
980         7
```

```
986        36
988        28
993         2
994         4
Name: Model, Length: 541, dtype: int64
0 = Fake news
1 = True news
```

[76]: `<function matplotlib.pyplot.show(close=None, block=None)>`

Fake news and true news

[77]:
```python
data = data_merge.drop(["Manufacturer","Prod.
 ↪year","Category","Cylinders",        "Wheel",        "Color",        "Airbags","Drive
 ↪wheels"],axis=1)
data.head(10)
```

[77]:
```
          ID  Levy    Model Leather interior Fuel type Engine volume  \
0   44020629     -     Golf               No    Diesel     2.0 Turbo
1   45784798     -   Sonata              Yes    Petrol           2.4
2   45800257     -    Tiida               No    Petrol           1.5
3   45797981   975     XC90              Yes    Petrol           3.2
4   45814303     -    Astra               No    Petrol           1.6
5   45812490   639      320              Yes    Petrol     2.0 Turbo
```

```
6   45772399   1646      GX 460                    Yes      Petrol            4.6
7   45766222   1172       E 350                    Yes      Diesel            3.5
8   45803362    891      Sonata                    Yes         LPG              2
9   45809004    645     Lacetti                    Yes      Petrol            1.8


       Mileage Gear box type   Doors   Price   class
0        0 km         Manual  02-Mar     NaN       0
1    26000 km      Tiptronic  04-May     NaN       0
2   168000 km      Automatic  04-May     NaN       0
3   143000 km      Automatic  04-May     NaN       0
4   200000 km         Manual  04-May     NaN       0
5    59733 km      Tiptronic  04-May     NaN       0
6   102907 km      Automatic  04-May     NaN       0
7    64496 km      Automatic  04-May     NaN       0
8   300234 km      Automatic  04-May     NaN       0
9   170000 km      Automatic  04-May     NaN       0
```

[78]:
```python
data=data.sample(frac=1)
data.head(10)
```

[78]:
```
             ID  Levy    Model Leather interior Fuel type Engine volume  \
5798   45740887  1079    Passat              Yes    Petrol     2.0 Turbo
16672  45768787   574   Insight              Yes    Hybrid           1.3
15993  45728197   645   CT 200h              Yes    Hybrid           1.8
3352   45813382   917       335              Yes    Petrol             3
14449  45814954     -    X-Trail              Yes       CNG             2
16005  45803379   765    Sonata              Yes       LPG             2
7516   45802658  1077  Santa FE              Yes    Diesel             2
6818   45730505  1099    REXTON              Yes    Diesel           2.7
14795  45797959   583   Elantra              Yes    Petrol           1.6
2123   45720276  1055     E 350              Yes    Petrol     3.5 Turbo


         Mileage Gear box type   Doors     Price  class
5798    31000 km      Tiptronic  04-May       NaN      0
16672  245070 km      Automatic  04-May     549.0      1
15993  228610 km      Automatic  04-May     470.0      1
3352   160696 km      Automatic  04-May       NaN      0
14449  125000 km      Automatic  04-May   10349.0      1
16005  510248 km      Automatic  04-May    9252.0      1
7516    35973 km      Automatic  04-May       NaN      0
6818   115321 km      Automatic  04-May       NaN      0
14795  175237 km      Automatic  04-May   16894.0      1
2123   140800 km      Automatic  04-May       NaN      0
```

[79]:
```python
data.isnull().sum()
```

```
[79]: ID                    0
      Levy                  0
      Model                 0
      Leather interior      0
      Fuel type             0
      Engine volume         0
      Mileage               0
      Gear box type         0
      Doors                 0
      Price              8208
      class                 0
      dtype: int64
```

```python
[80]: def filtering(data):
        text=data.lower()
        text=re.sub('\[.*?\]','',text)
        text=re.sub("\\W","",text)
        text=re.sub('https?://\s+|www\.S+','',text)
        text=re.sub('<.*?>+','',text)
        text=re.sub('[%s]'% re.escape(string.punctuation),'',text)
        text=re.sub('\w*\d\w*','',text)
        return text
```

```python
[81]: data["Model"]= data["Model"].apply(filtering)
      data.head(10)
```

```
[81]:              ID  Levy    Model Leather interior Fuel type Engine volume  \
      5798   45740887  1079   passat              Yes    Petrol     2.0 Turbo
      16672  45768787   574  insight              Yes    Hybrid           1.3
      15993  45728197   645                       Yes    Hybrid           1.8
      3352   45813382   917                       Yes    Petrol             3
      14449  45814954     -   xtrail              Yes       CNG             2
      16005  45803379   765   sonata              Yes       LPG             2
      7516   45802658  1077   santafe             Yes    Diesel             2
      6818   45730505  1099   rexton              Yes    Diesel           2.7
      14795  45797959   583  elantra              Yes    Petrol           1.6
      2123   45720276  1055                       Yes    Petrol     3.5 Turbo

                Mileage Gear box type   Doors    Price  class
      5798    31000 km     Tiptronic  04-May      NaN      0
      16672  245070 km     Automatic  04-May    549.0      1
      15993  228610 km     Automatic  04-May    470.0      1
      3352   160696 km     Automatic  04-May      NaN      0
      14449  125000 km     Automatic  04-May  10349.0      1
      16005  510248 km     Automatic  04-May   9252.0      1
      7516    35973 km     Automatic  04-May      NaN      0
      6818   115321 km     Automatic  04-May      NaN      0
```

```
14795   175237 km      Automatic   04-May  16894.0      1
2123    140800 km      Automatic   04-May     NaN       0
```

[82]:
```python
x=data["Levy"]
y=data["Model"]
```