

Міністерство освіти і науки України

Національний університет «Запорізька політехніка»

кафедра програмних засобів

ЗВІТ

з лабораторної роботи № 2

з дисципліни «Інтелектуальний аналіз даних» на тему:

«ОСНОВИ ПОПЕРЕДНЬОЇ ОБРОБКИ ДАНИХ»

Виконав:

ст. гр. КНТ-113сп

Іван ЩЕДРОВСЬКИЙ

Прийняв:

зав. лабораторії

Максим АНДРЕЄВ

2025

1 Мета роботи

Ознайомитися та отримати навички роботи з програмою WEKA та бібліотеками мови програмування Python для проведення аналізу даних. На практиці вивчити методи попередньої обробки даних для задач інтелектуального аналізу даних

2 Завдання до лабораторної роботи

2.1. Обрати в додатку В вибірку для аналізу.

2.2. Виконати попередню обробку даних за допомогою програми WEKA:

- завантажити вибірку;

- описати яке практичне завдання вирішується;

- описати всі атрибути, які характеризують екземпляри вибірки, та їх типи даних;

- визначити, який атрибут є цільовим, які значення він приймає та скільки екземплярів кожного класу в вибірці;

- визначити чи є екземпляри з відсутніми значеннями. Якщо такі екземпляри є, то встановити їм відповідні значення використовуючи одну з наведених стратегій поведінки при роботі з пропущеними значеннями;

- визначити чи є викиди даних;

- подати дані в графічному вигляді.

3. Виконати попередню обробку даних за допомогою засобів мови програмування Python:

- конвертувати вибрану для аналізу вибірку з ARFF до CSV формату;

- використовуючи бібліотеку pandas виконати попередню обробку даних (завантажити файл, вивести на екран частину вибірки, для пропущених значень використати одну з наведених стратегій поведінки);

- використовуючи бібліотеку matplotlib побудувати необхідні графіки для подання даних в графічному вигляді (типи графіків: pie, hist, scatter, bar);

3 Виконання лабораторної роботи

Згідно додатку В методичних вказівок було обрано набір даних з варіантом 14, $29 \bmod 16 + 1 = 14$, який називається vote.arff. Повна назва цієї вибірки «1984 United States Congressional Voting Records Database»

Вибірка має 16 boolean полів та одне поле для класу, всього 17:

- class name – назва класу, демократи або республіканці
- handicapped-infants – Законопроект про фінансування та підтримку програм для дітей з інвалідністю. Y – фінансування потрібно
- water-project-cost-sharing – Місцева влада повинна також платити за проекти водної інфраструктури, наприклад, побудову дамб і так далі. Y – повинна
- adoption-of-the-budget-resolution – Підтримка загального державного бюджету
- physician-fee-freeze – голосування щодо заморожування гонорарів лікарів у державних медичних програмах
- el-salvador-aid — голосування щодо надання фінансової або військової допомоги Сальвадору.
- religious-groups-in-schools — голосування щодо дозволу діяльності релігійних груп у державних школах.
- anti-satellite-test-ban — голосування щодо заборони тестування протисупутникової зброї.
- aid-to-nicaraguan-contras — голосування щодо підтримки антикомуністичних повстанців у Нікарагуа.
- mx-missile — голосування щодо розгортання міжконтинентальної балістичної ракети MX.
- immigration — голосування щодо змін у законодавстві про імміграцію.
- synfuels-corporation-cutback — голосування щодо скорочення фінансування державної корпорації синтетичного палива.

- education-spending — голосування щодо рівня державних витрат на освіту.
- superfund-right-to-sue — голосування щодо права громадян подавати позови за екологічну шкоду.
- crime — голосування щодо посилення або зміни заходів боротьби зі злочинністю.
- duty-free-exports — голосування щодо надання податкових пільг на безмитний експорт.
- export-administration-act-south-africa — голосування щодо обмеження експорту до Південної Африки у зв'язку з апартеїдом.

Практичного завдання в аналізі цієї вибірки так якого немає

Дані було завантажено в Weka, це показано на рисунку 3.1

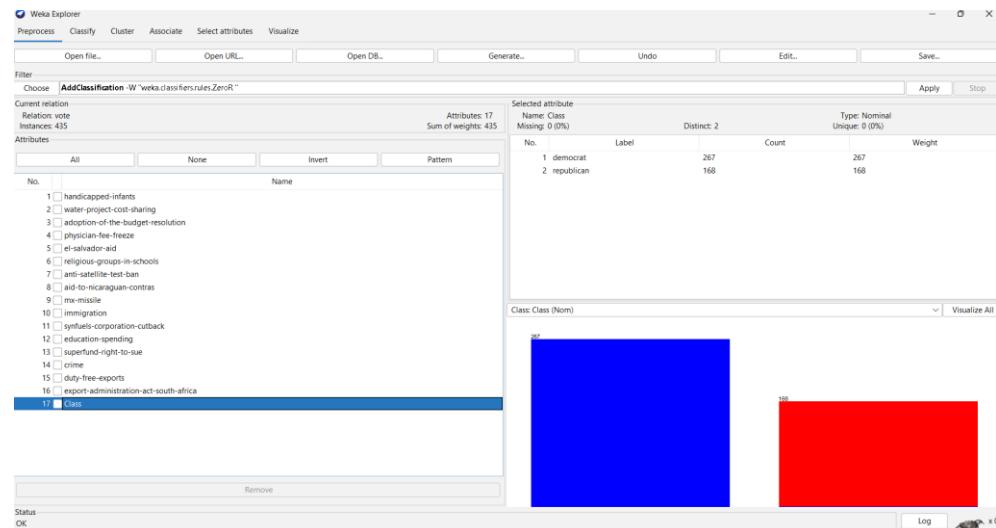


Рисунок 3.1 – Завантажені дані в Weka

Цільовим атрибутом є поле Class name, яке означає чи це демократ, чи республіканець

На рисунку 3.2 показана візуалізація всіх атрибутів з вибірки

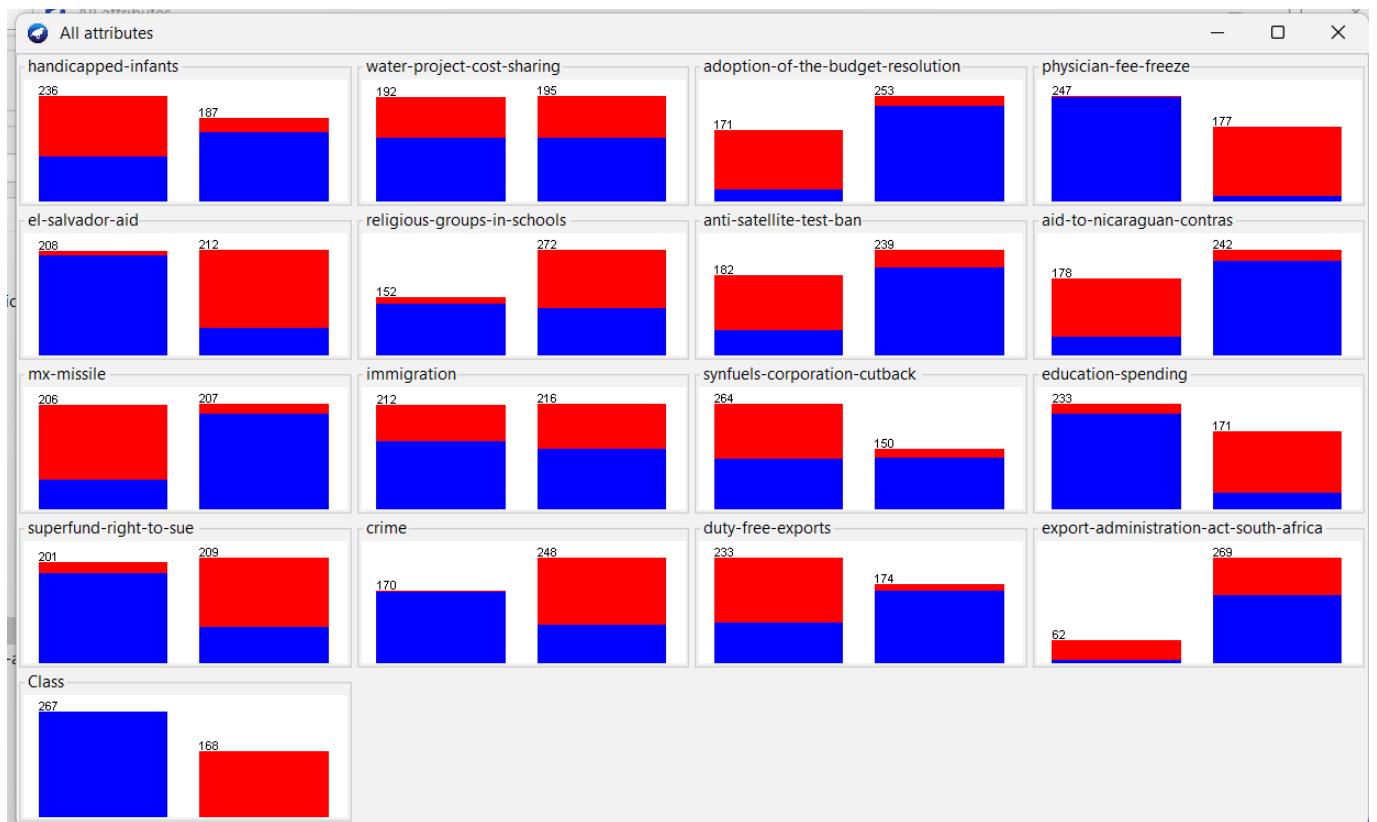


Рисунок 3.2 – Візуалізація всіх атрибутів

В вибірці було переглянуто всі дані та вручну видалено два повністю пусті рядки

Далі всі відсутні значення було заповнено на основі середніх по класу через використання ReplaceMissingValues. Це показано на рисунку 3.3

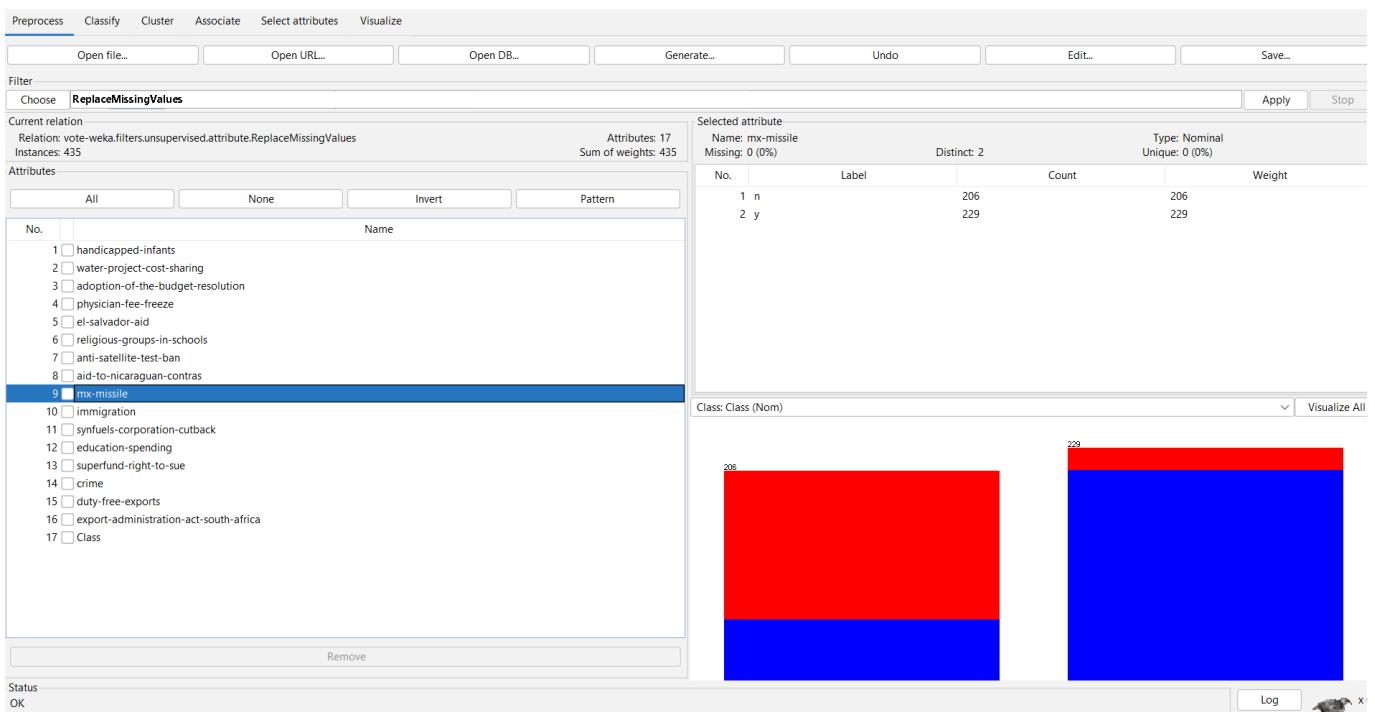


Рисунок 3.3 – Замінені відсутні значення

Також тепер можна виконати візуалізацію даних та переглянути відмінності, це показано на рисунку 3.4

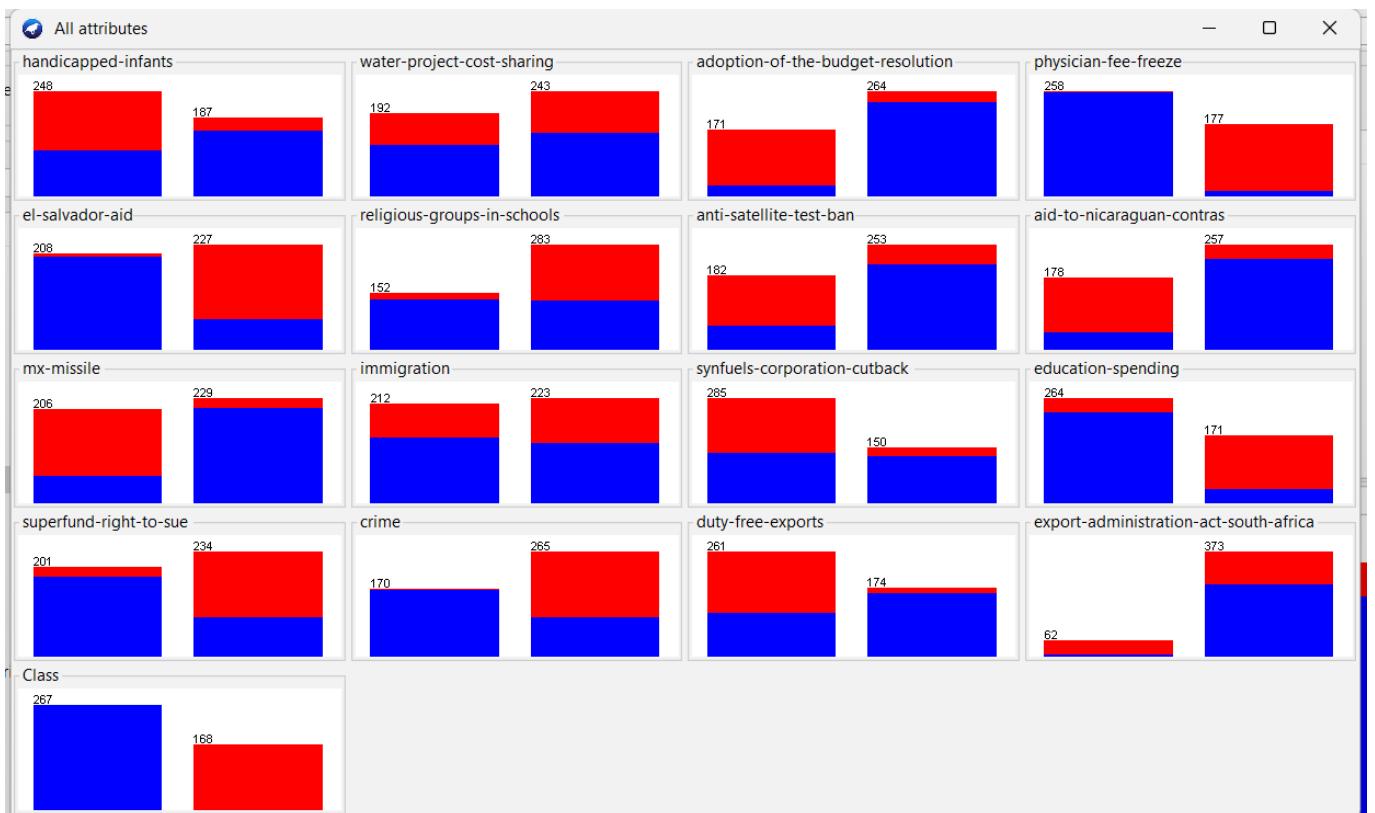


Рисунок 3.4 – Візуалізація даних після заміни

Далі було виконано відбір атрибутів. Це показано на рисунку 3.5

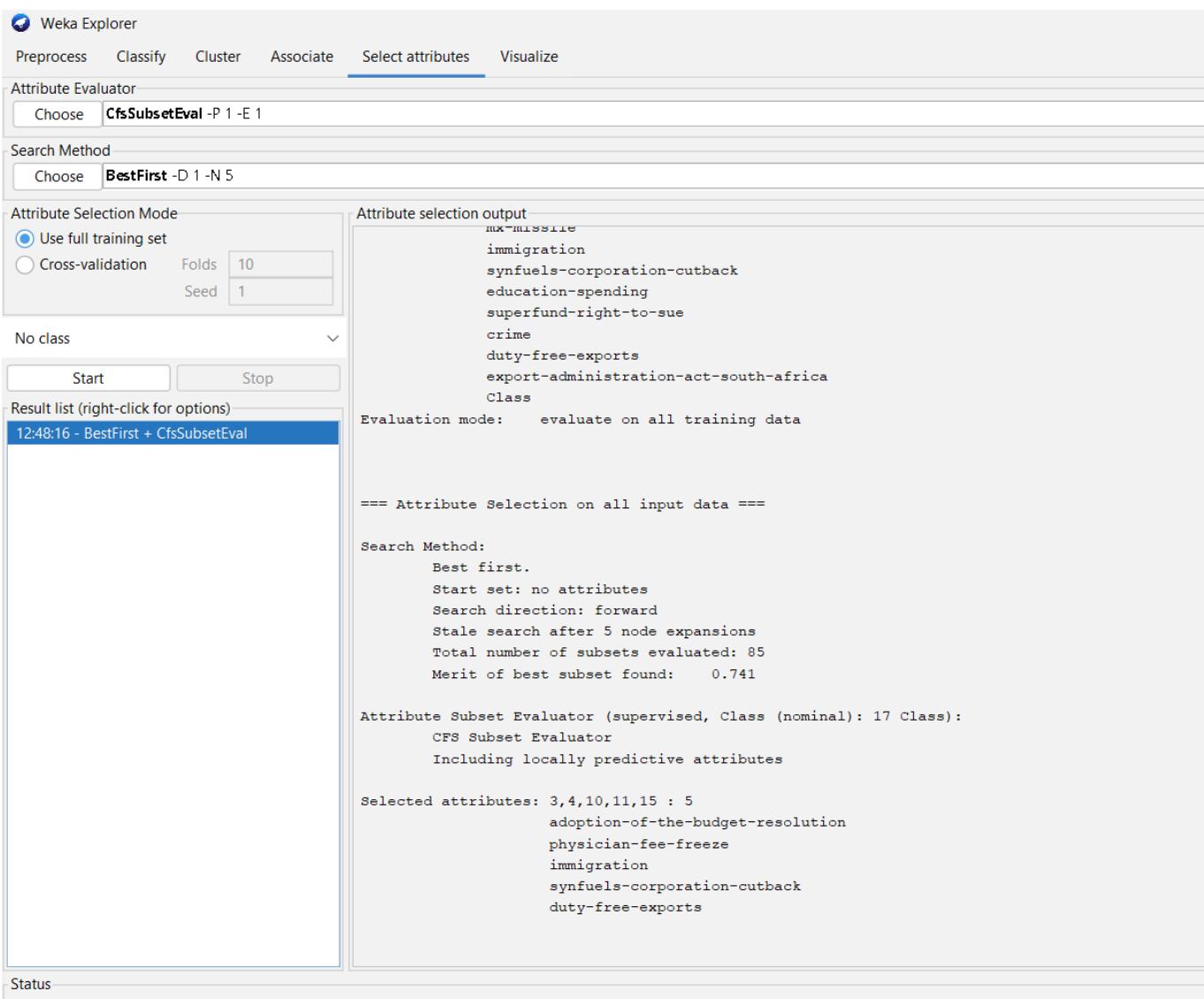


Рисунок 3.5 – Відбір атрибутів

Після відбору атрибутів всі інші були видалені. Це показано на рисунку 3.6

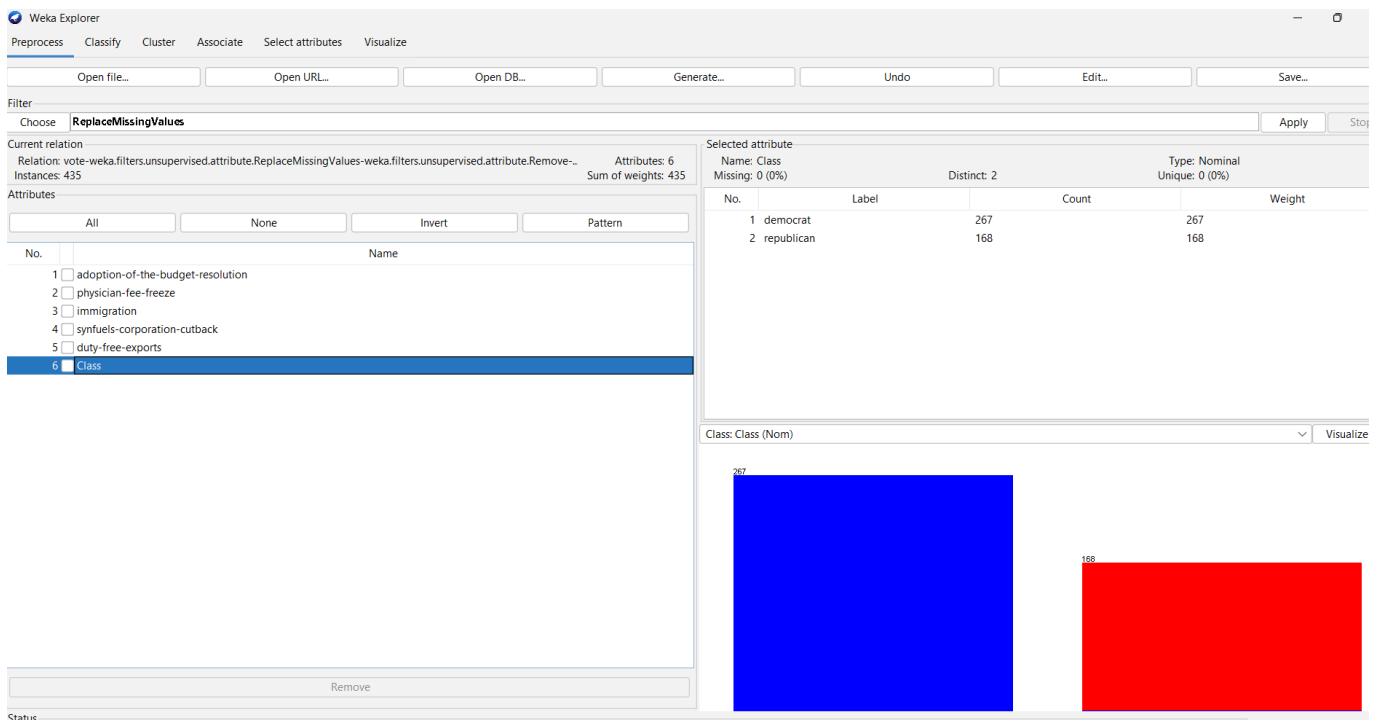


Рисунок 3.6 – Видалені атрибути

I виконана візуалізація даних, це показано на рисунку 3.7

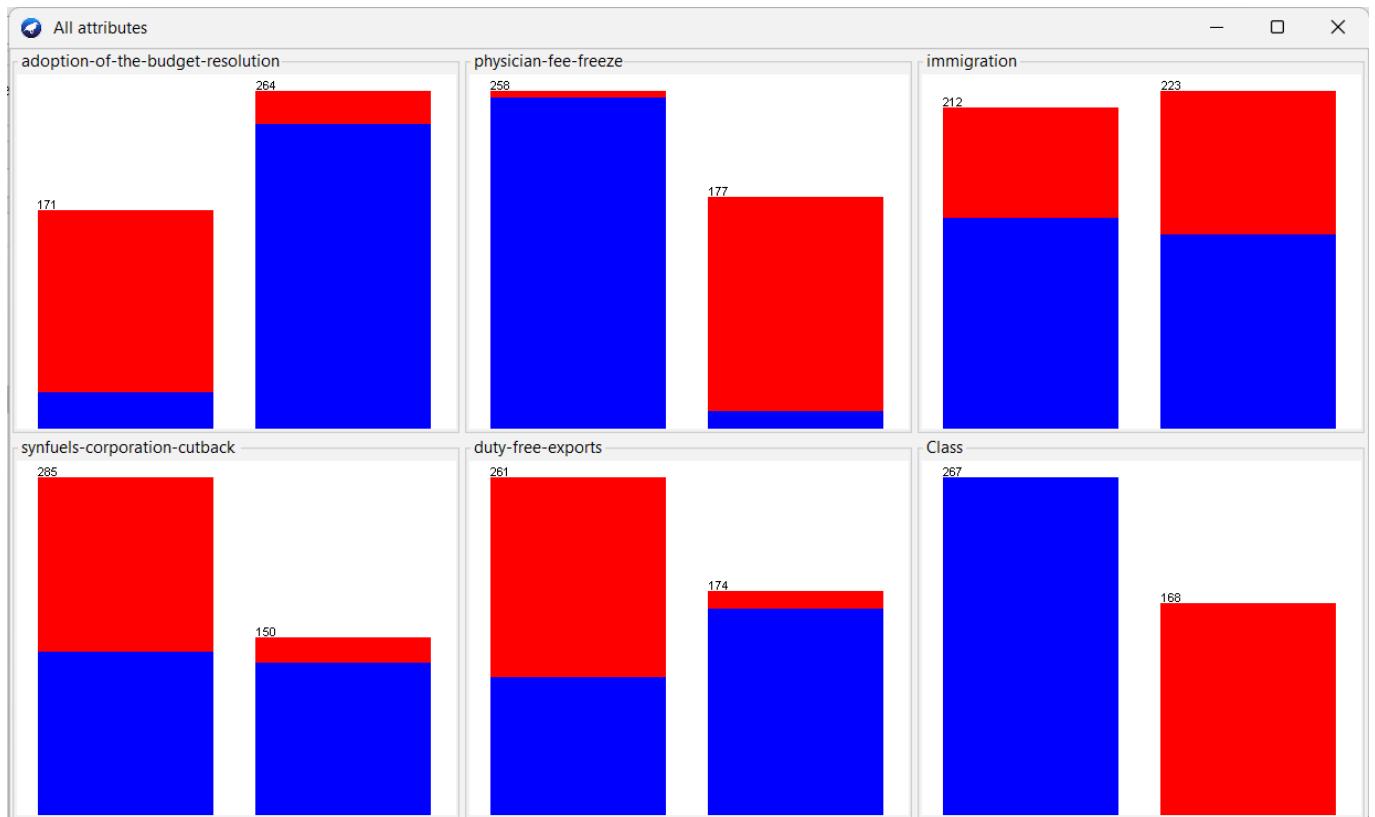


Рисунок 3.7 – Візуалізація вибірки

На цьому етапі робота з Weka завершена та вибірка готова до аналізу

Далі, для виконання обробки даних в python початкову вибірку було переведено в формат csv через використання Weka. Це показано на рисунку 3.8. Важливо зауважити, що дані, показані на рисунку є візуалізацією PyCharm з врахуванням transpose матриці даних

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	2
C1	handicapped-infants	n	n	?	n	y	n	n	n	y	n	n	n	y	n	n	y	y	n	y	
C2	water-project-cost-sharing	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	?	y	y	
C3	adoption-of-the-budget-resol...	n	n	y	y	y	n	n	n	y	n	n	y	y	n	n	y	y	n	y	
C4	physician-fee-freeze	y	y	?	n	n	y	y	y	n	y	y	n	n	y	y	n	n	y	n	
C5	el-salvador-aid	y	y	y	?	y	y	y	y	y	n	y	y	n	n	y	y	n	n	y	n
C6	religious-groups-in-schools	y	y	y	y	y	y	y	y	y	n	n	y	y	y	y	y	y	n	y	n
C7	anti-satellite-test-ban	n	n	n	n	n	n	n	n	y	n	n	y	y	n	n	n	y	n	y	n
C8	aid-to-nicaraguan-contras	n	n	n	n	n	n	n	n	y	n	n	y	y	n	n	y	y	n	y	n
C9	mx-missile	n	n	n	n	n	n	n	n	y	n	n	y	?	n	n	?	y	n	y	n
C10	immigration	y	n	n	n	n	n	n	n	n	n	n	n	y	n	y	y	n	n	n	n
C11	synfuels-corporation-cutback	?	n	y	y	y	n	n	n	n	?	y	n	y	n	n	y	n	n	y	n
C12	education-spending	y	y	n	n	?	n	n	n	y	n	?	?	n	?	y	y	y	n	?	n
C13	superfund-right-to-sue	y	y	y	y	y	?	y	y	n	y	y	y	n	?	y	?	y	y	y	n

Рисунок 3.8 – Збереження початкової вибірки в форматі csv

Для Python було встановлено бібліотеку pandas, це показано на рисунку 3.9

```
(.venv) PS C:\home\university-4\data-mining\code> pip install pandas
Collecting pandas
  Downloading pandas-2.3.3-cp311-cp311-win_amd64.whl.metadata (19 kB)
Collecting numpy>=1.23.2 (from pandas)
  Downloading numpy-2.3.5-cp311-cp311-win_amd64.whl.metadata (60 kB)
Collecting python-dateutil>=2.8.2 (from pandas)
  Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)
Collecting pytz>=2020.1 (from pandas)
  Downloading pytz-2025.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)
  Downloading tzdata-2025.3-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting six>=1.5 (from python-dateutil>=2.8.2->pandas)
  Downloading six-1.17.0-py2.py3-none-any.whl.metadata (1.7 kB)
Downloading pandas-2.3.3-cp311-cp311-win_amd64.whl (11.3 MB)
----- 11.3/11.3 MB 7.8 MB/s eta 0:00:00
Downloading numpy-2.3.5-cp311-cp311-win_amd64.whl (13.1 MB)
----- 13.1/13.1 MB 7.7 MB/s eta 0:00:00
```

Рисунок 3.9 – Встановлення pandas

Також було встановлено бібліотеку matplotlib, це показано на рисунку 3.10

```
(.venv) PS C:\home\university-4\data-mining\code> pip install matplotlib
Collecting matplotlib
  Downloading matplotlib-3.10.8-cp311-cp311-win_amd64.whl.metadata (52 kB)
Collecting contourpy>=1.0.1 (from matplotlib)
  Downloading contourpy-1.3.3-cp311-cp311-win_amd64.whl.metadata (5.5 kB)
Collecting cycler>=0.10 (from matplotlib)
  Downloading cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)
Collecting fonttools>=4.22.0 (from matplotlib)
  Downloading fonttools-4.61.1-cp311-cp311-win_amd64.whl.metadata (116 kB)
Collecting kiwisolver>=1.3.1 (from matplotlib)
  Downloading kiwisolver-1.4.9-cp311-cp311-win_amd64.whl.metadata (6.4 kB)
Requirement already satisfied: numpy>=1.23 in c:\home\university-4\data-mining\code\.venv
.3.5)
Collecting packaging>=20.0 (from matplotlib)
  Downloading packaging-25.0-py3-none-any.whl.metadata (3.3 kB)
Collecting pillow>=8 (from matplotlib)
  Downloading pillow-12.0.0-cp311-cp311-win_amd64.whl.metadata (9.0 kB)
Collecting pyparsing>=3 (from matplotlib)
  Downloading pyparsing-3.2.5-py3-none-any.whl.metadata (5.0 kB)
Requirement already satisfied: python-dateutil>=2.7 in c:\home\university-4\data-mining\c
```

Рисунок 3.10 – Встановлення matplotlib

Далі на Python з використанням pandas було написано програму, яка читає дані з csv, перетворює їх в Boolean та заповнює всі пропущені значення на значення, яке зустрічається в класі найбільше

Для цього прикладу можливо це не є найкращим підходом, оскільки один голос може вирішувати багато що, але це краще, а ніж повне видалення даних

На рисунках 3.11 та 3.12 показано виконання програми

```
C:\home\university-4\data-mining\code\.venv\Scripts\python.exe C:
\home\university-4\data-mining\code\lb2.py
handicapped-infants          object
water-project-cost-sharing    object
adoption-of-the-budget-resolution object
physician-fee-freeze          object
el-salvador-aid               object
religious-groups-in-schools   object
anti-satellite-test-ban       object
aid-to-nicaraguan-contras     object
mx-missile                     object
immigration                     object
synfuels-corporation-cutback  object
education-spending              object
superfund-right-to-sue         object
crime                           object
duty-free-exports               object
export-administration-act-south-africa object
Class                           object
dtype: object

-- Data is:

   handicapped-infants ... Class
0      False ... republican
1      False ... republican
2      NaN ... democrat
3      False ... democrat
4      True ... democrat
...
430     False ... republican
431     False ... democrat
432     False ... republican
433     False ... republican
434     False ... republican

[435 rows x 17 columns]

-- Data info is:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 435 entries, 0 to 434
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   handicapped-infants  423 non-null   object 
 1   water-project-cost-sharing 387 non-null   object 
 2   adoption-of-the-budget-resolution 424 non-null   object 
 3   physician-fee-freeze  424 non-null   object 
 4   el-salvador-aid     420 non-null   object 
 5   religious-groups-in-schools 424 non-null   object 
 6   anti-satellite-test-ban 421 non-null   object 
 7   aid-to-nicaraguan-contras 420 non-null   object 
 8   mx-missile          413 non-null   object 
 9   immigration          428 non-null   object 
 10  synfuels-corporation-cutback 414 non-null   object 
 11  education-spending    404 non-null   object 
 12  superfund-right-to-sue 410 non-null   object 
 13  crime                418 non-null   object 
 14  duty-free-exports     407 non-null   object 
 15  export-administration-act-south-africa 331 non-null   object 
 16  Class                435 non-null   object 
dtypes: object(17)
memory usage: 57.9+ KB
None

-- Data of handicapped-infants is:

0      False
1      False
2      NaN
3      False
4      True
...
430     False
431     False
432     False
433     False
434     False
Name: handicapped-infants, Length: 435, dtype: object
```

Рисунок 3.11 – Виконання програми 1

```
-- Data of 1 row is:
handicapped-infants      False
water-project-cost-sharing True
adoption-of-the-budget-resolution False
physician-fee-freeze      True
el-salvador-aid          True
religious-groups-in-schools True
anti-satellite-test-ban  False
aid-to-nicaraguan-contras False
mx-missile                False
immigration               False
synfuels-corporation-cutback False
education-spending        True
superfund-right-to-sue    True
crime                     True
duty-free-exports         False
export-administration-act-south-africa NaN
Class                      republican
Name: 1, dtype: object

-- Sum of null fields:
handicapped-infants      12
water-project-cost-sharing 48
adoption-of-the-budget-resolution 11
physician-fee-freeze      11
el-salvador-aid          15
religious-groups-in-schools 11
anti-satellite-test-ban  14
aid-to-nicaraguan-contras 15
mx-missile                22
immigration               7
synfuels-corporation-cutback 21
education-spending        31
superfund-right-to-sue    25
crime                     17
duty-free-exports         28
export-administration-act-south-africa 104
Class                      0
dtype: int64

-- Sum of null fields:
handicapped-infants      0
water-project-cost-sharing 0
adoption-of-the-budget-resolution 0
physician-fee-freeze      0
el-salvador-aid          0
religious-groups-in-schools 0
anti-satellite-test-ban  0
aid-to-nicaraguan-contras 0
mx-missile                0
immigration               0
synfuels-corporation-cutback 0
education-spending        0
superfund-right-to-sue    0
crime                     0
duty-free-exports         0
export-administration-act-south-africa 0
Class                      0
dtype: int64
      handicapped-infants ... Class
0           False ... republican
1           False ... republican
2            True ... democrat
3           False ... democrat
4            True ... democrat
...
430          False ... republican
431          False ... democrat
432          False ... republican
433          False ... republican
434          False ... republican
[435 rows x 17 columns]
handicapped-infants      bool
water-project-cost-sharing bool
adoption-of-the-budget-resolution bool
physician-fee-freeze      bool
el-salvador-aid          bool
religious-groups-in-schools bool
anti-satellite-test-ban  bool
aid-to-nicaraguan-contras bool
mx-missile                bool
immigration               bool
synfuels-corporation-cutback bool
education-spending        bool
superfund-right-to-sue    bool
crime                     bool
duty-free-exports         bool
export-administration-act-south-africa bool
Class                      category
dtype: object
```

Рисунок 3.12 – Виконання програми 2

На рисунку 3.13 показаний вихідний код застосунку

```

import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("./vote-csv.csv", na_values=["?"])
data = data.replace({'y': True, 'n': False})

print(data.dtypes)
print("\n -- Data is:\n")
print(data)

print("\n -- Data info is:\n")
print(data.info())

print("\n -- Data of handicapped-infants is:\n")
print(data["handicapped-infants"])

print("\n -- Data of 1 row is:\n")
print(data.loc[1])

print("\n -- Sum of null fields:\n")
print(data.isnull().sum())

data = data.fillna(data.groupby("Class").transform(lambda x: x.mode()[0]))

for col in data.columns.drop("Class"):
    data[col] = data[col].astype("bool")
data["Class"] = data["Class"].astype("category")

print("\n -- Sum of null fields:\n")
print(data.isnull().sum())
print(data)
print(data.dtypes)

data["Class"].value_counts().plot(kind = "bar", figsize=(5, 5))
plt.show()

fig, axes = plt.subplots(nrows=4, ncols=4, figsize=(20, 20))
xa=0
ya=0

for col in data.columns.drop("Class"):
    plt.title(col)
    data[col].value_counts().plot(
        kind='bar',
        color=['green', 'red'],
        ax=axes[xa, ya],
    )
    for i, count in enumerate(data[col].value_counts()):
        axes[xa, ya].text(
            i,
            count + 0.1,
            str(count),
            ha='center',
            va='bottom',
            fontsize=9,
            color='black'
        )
    if xa < 3:
        xa += 2
    else:
        ya += 1
        xa=0

plt.tight_layout()
plt.show()

save = [
    "adoption-of-the-budget-resolution",
    "physician-fee-freeze",
    "immigration",
    "synfuels-corporation-cutback",
    "duty-free-exports",
]
]

data = data.replace({True: "y", False: "n"})
data.to_csv("./vote-python-raw.csv", index=False)

for col in data.columns.drop("Class"):
    if (col not in save):
        data = data.drop([col], axis=1)

data.to_csv("./vote-python.csv", index=False)

```

Рисунок 3.13 – Код застосунку

Також застосунок використовує matplotlib для візуалізації даних. Два графіки показані на рисунках 3.14 та 3.15

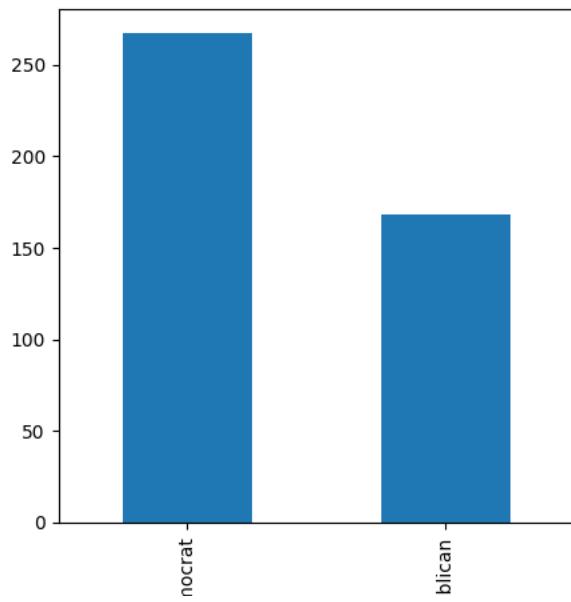


Рисунок 3.14 – Графік кількості демократів до республіканців

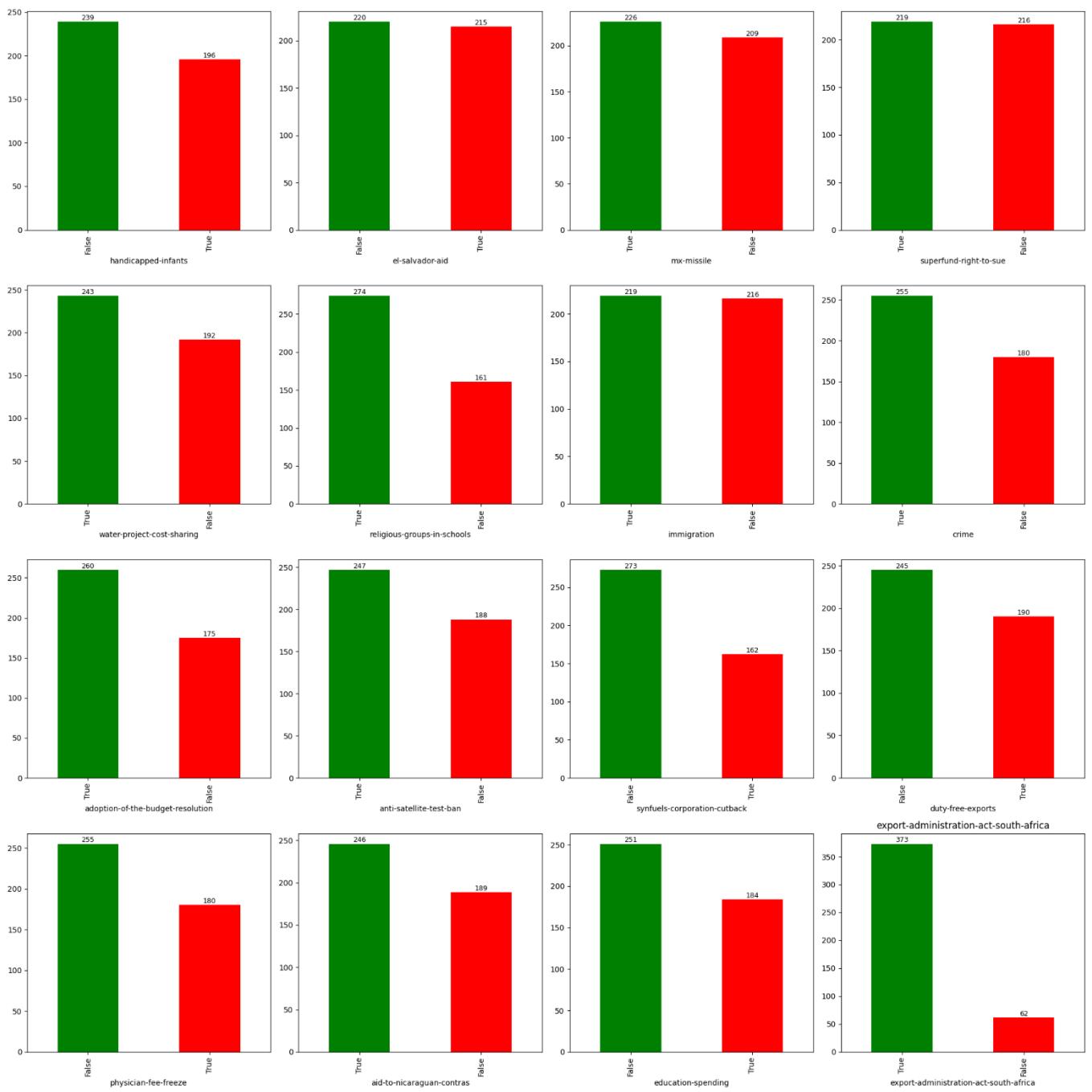


Рисунок 3.15 – Графік по кожному атрибуту

Також було виконано порівняння версії з python з початковими даними, а також версії python з видаленими стовпцями з waka, це показано на рисунках 3.16 та 3.17 відповідно

vote-python-raw.csv

```
handicapped-infants,water-project-cost-sharing,  
adoption-of-the-budget-resolution,physician-fee-freeze,  
el-salvador-aid,religious-groups-in-schools,  
anti-satellite-test-ban,aid-to-nicaraguan-contras,  
mx-missile,immigration,synfuels-corporation-cutback,  
education-spending,superfund-right-to-sue,crime,  
duty-free-exports,export-administration-act-south-africa,  
Class  
n,y,n,y,y,y,n,n,n,y,n,y,y,y,n,y,republican  
n,y,n,y,y,y,n,n,n,n,y,y,y,n,y,republican  
y,y,y,n,y,y,n,n,n,y,n,y,y,n,n,democrat  
n,y,y,n,n,y,n,n,n,n,y,n,y,n,n,y,democrat  
y,y,y,n,y,y,n,n,n,n,y,n,y,y,y,democrat  
n,y,y,n,y,y,n,n,n,n,n,y,y,y,y,democrat  
n,y,n,y,y,y,n,n,n,n,n,y,y,y,y,democrat  
n,y,n,y,y,y,n,n,n,n,n,y,y,n,y,republican  
n,y,n,y,y,y,n,n,n,n,n,y,y,n,y,republican  
y,y,y,n,n,y,y,y,n,n,n,n,y,y,y,y,democrat  
n,y,n,y,y,y,n,n,n,n,n,y,y,y,n,n,replican  
n,y,n,y,y,y,n,n,n,n,n,y,y,y,n,n,replican  
n,y,n,y,y,y,n,n,n,n,y,y,y,y,n,y,republican  
n,y,y,n,n,y,y,y,y,n,n,y,n,y,y,democrat  
y,y,y,n,n,y,y,y,y,y,n,n,y,y,y,democrat  
n,y,n,y,y,y,n,n,n,n,y,y,y,y,n,y,republican  
n,y,n,y,y,y,y,n,n,n,y,y,y,y,n,y,republican  
n,y,n,y,n,n,y,y,y,y,y,n,n,y,y,democrat  
y,y,y,n,n,y,y,y,y,y,y,n,n,y,y,y,democrat  
n,y,n,y,y,y,n,n,n,n,y,y,y,y,n,y,republican  
n,y,n,y,y,y,y,n,n,y,y,y,y,n,y,republican  
n,y,n,y,n,n,y,y,y,y,y,n,n,y,y,democrat  
y,y,y,n,n,y,y,y,y,y,y,n,n,y,y,y,democrat  
n,y,n,y,y,y,n,n,n,n,y,y,y,y,n,y,republican
```

vote-csv.csv

1	1	1	handicapped-infants,water-project-cost-sharing, adoption-of-the-budget-resolution,physician-fee-freeze, el-salvador-aid,religious-groups-in-schools, anti-satellite-test-ban,aid-to-nicaraguan-contras, mx-missile,immigration,synfuels-corporation-cutback, education-spending,superfund-right-to-sue,crime, duty-free-exports,export-administration-act-south-africa, Class
2	2	2 <<	n,y,n,y,y,y,n,n,n,y,?,y,y,n,y,republican
3	3	3	n,y,n,y,y,y,n,n,n,n,y,y,n,?,republican
4	4	4	2,y,y,?,y,y,n,n,n,n,y,n,y,y,n,n,democrat
5	5	5	n,y,y,n,?,y,y,n,n,n,n,y,n,y,n,n,y,democrat
6	6	6	y,y,y,n,y,y,n,n,n,n,y,?,y,y,y,y,democrat
7	7	7	n,y,y,n,y,y,n,n,n,n,n,y,y,y,y,democrat
8	8	8 <<	n,y,n,y,y,y,y,n,n,n,n,n,?,y,y,y,democrat
9	9	9	n,y,n,y,y,y,y,n,n,n,n,n,y,y,?,y,republican
10	10	10	n,y,n,y,y,y,y,n,n,n,n,y,y,y,n,y,republican
11	11	11 <<	y,y,y,n,n,n,y,y,y,n,n,n,n,?,?,democrat
12	12	12	n,y,n,y,y,y,n,n,n,n,n,y,?,y,y,n,n,republican
13	13	13	n,y,n,y,y,y,y,n,n,n,n,y,?,y,y,?,?,,republican
14	14	14	n,y,y,n,n,n,y,y,y,y,n,n,n,y,n,?,?,,democrat
15	15	15	y,y,y,n,n,y,y,y,?,y,y,?,n,y,?,democrat
16	16	16	n,y,n,y,y,y,y,n,n,n,n,y,?,?,,n,?,republican
17	17	17	n,y,n,y,y,y,y,n,n,n,y,n,y,?,?,,n,?,republican
18	18	18	y,n,y,n,n,y,y,y,?,y,y,?,n,n,y,democrat
19	19	19	2,y,y,n,n,n,y,y,y,y,y,n,n,y,n,y,y,democrat
20	20	20	n,y,n,y,y,y,y,n,n,n,n,y,?,y,y,n,n,republican

Рисунок 3.16 – Порівняння даних після перетворення з початковими

vote-python.csv	vote-weka.csv
n,y,y,n,n,republican	2 2 n,y,y,n,n,republican
n,y,n,n,n,republican	3 3 n,y,n,n,n,republican
y,n,n,y,n,democrat	4 4 y,n,n,y,n,democrat
y,n,n,y,n,democrat	5 5 y,n,n,y,n,democrat
y,n,n,y,y,democrat	6 6 y,n,n,y,y,democrat
y,n,n,n,y,democrat	7 7 y,n,n,n,y,democrat
n,y,n,n,y,democrat	8 8 n,y,n,n,y,democrat
n,y,n,n,n,republican	9 9 n,y,n,n,n,republican
n,y,n,n,n,republican	10 10 n,y,n,n,n,republican
y,n,n,n,y,democrat	» 11 11 « y,n,n,n,n,democrat
n,y,n,n,n,republican	12 12 n,y,n,n,n,republican
n,y,n,y,n,republican	13 13 n,y,n,y,n,republican
y,n,n,n,y,democrat	» 14 14 « y,n,n,n,n,democrat
y,n,Y,y,y,democrat	15 15 y,n,Y,y,y,democrat
n,y,n,n,n,republican	16 16 n,y,n,n,n,republican
n,y,y,n,n,republican	17 17 n,y,y,n,n,republican
y,n,Y,y,n,democrat	18 18 y,n,Y,y,n,democrat
y,n,n,n,y,democrat	19 19 y,n,n,n,y,democrat
n,y,n,n,n,republican	20 20 n,y,n,n,n,republican
y,n,n,y,y,democrat	21 21 y,n,n,y,y,democrat
y,n,n,n,y,democrat	22 22 y,n,n,n,y,democrat
y,n,n,n,y,democrat	23 23 y,n,n,n,y,democrat
y,n,n,n,y,democrat	24 24 y,n,n,n,y,democrat
y,n,n,n,y,democrat	25 25 y,n,n,n,y,democrat
y,n,n,n,y,democrat	26 26 y,n,n,n,y,democrat
y,n,y,n,y,democrat	27 27 y,n,y,n,y,democrat
y,n,n,y,y,democrat	28 28 y,n,n,y,y,democrat

Рисунок 3.17 – Порівняння даних з python та weka

Як можна побачити, початкові дані були сильно змінені, аж 105 ліній даних. Дані з python також відрізняються від даних з Weka, але не настільки сильно, на 28 відмінностей. Ці відмінності могли бути викликані тим, що при використанні python не видалялись пусті лінії, а також тим, що pandas працює по іншому

4 Висновки

Я ознайомився та отримав навички роботи з програмою WEKA та бібліотеками мови програмування Python для проведення аналізу даних. На практиці вивчив методи попередньої обробки даних для задач інтелектуального аналізу даних