Introduction to Data Mining
Ali Can Mutlu Individual Report

Individual Contributions Report

1.Introduction

As group pollution solvers, we tried to tackle the issue of the effects of energy consumption by specific type of energy on the deaths caused by pollution. While we were building up the project the first obstacle that we had to overcome was to find a problem to solve. So for our first meeting we came up with topic proposals of our own. After non-heated discussions we decided to ramble with a topic that interested us all, which was the topic of energy consumption and pollution. One question that we had regarding this topic was the fact that we were not sure if we were allowed to use multiple data sets. We talked with the professor and the TA which ended with us, using multiple datasets.

2.Description of your individual work

One of the most important aspects of our education in this Master's program is the fact that we are all international students. Being in a country that one's not familiar with is already hard but getting an education in your 2nd language and trying to communicate in that language with your peers is even harder. For this reason I took on the role of being the planner and the communicator of the group. After talking with Ian and Zhong Yang, we quickly made a decision to which social messaging platform to use and then we started communicating through that platform(WhatsApp).

My main role in the project was to make the project as presentable as possible due to my business school background. So we decided that I should do the bulk of presentation preparation and checking of the grammar in written documents. On top of that I was involved in GUI code and some of the data processing coding. We all did some sort of data preprocessing in our own, yet we did not put all of it in the final code. As this was merely exploratory data analysis on our own in a way.

3.Describe the Portion of the Work

As I don't come from a coding background. Most of the topics that we went through in the class was especially difficult for me. While the idea of GUI made me excited, building it was very hard. My teammate Zhong Yang advised me to use the Tinker and estimator packages to build the GUI. We thought that building it this way would allow the group to represent the linear regression models that we have in the best way.

```
plt.figure(figsize=(16,20))- For adjusting the canvas to include all portions
of the heatmap.


from tkinter import * ##Importing the packages
from estimator import estimator ## Importing the packages
```

```
from PIL import Image ## Importing the packages
import os ## Importing the packages
from pathlib import Path ## Importing the packages

est=estimator( )
cwd=os.getcwd()

pngs=Path(cwd).glob('*.png') ## showing path to get the png
png_dict={os.path.basename(i):i for i in pngs}


print(png_dict)

def est_button():  ##Defining button class
    t.delete('1.0',END)
    para=entry.get()
    ans=est.pred(para)
    t.insert('1.0',ans)
    return

def show_list(): ##Defining show list class
    t.delete('1.0',END)
    list1=[i for i in png_dict.keys()]
    for x in range(len(list1)):
        a=x+1
        n=str(float(1+x))
        content = list1[x]
        t.insert(n,str(a)+': '+content+'\n')

    return

def show_png(): #Defining showlist class
    list1=[i for i in png_dict.keys()]
    ind = entry.get()

    name=list1[int(int(ind)-1)]

    png = PhotoImage(file=name)
    canvas.create_image(0,0, anchor=NW, image=png)
    canvas.png=png
    return

root = Tk()

root.geometry('1200x700') #Adjusting the Gui page size for estimator section
root.title('Pollution Estimator')
content1= Label(root, text = 'Input value:', font=(50), width = 10)
entry= Entry(root)
entry.grid(row=1,column=1)
content1.grid(row=1,column=0)



estimate_button = Button(root, text='estimate',comman=est_button)
estimate_button.grid(row=1,column=2)
pre_button=Button(root,text='Show Images List',command=show_list)
next_button=Button(root,text='Show Image',command=show_png)
```

```
pre_button.grid(row=1,column=3)
next_button.grid(row=1,column=4)



t = Text(root,height=30,width=50) ## Setting the size of the text
t.config(font=30)
t.grid(row=0,column=0,columnspan=3)
canvas=Canvas(root,width=640,height=500) ## Setting the size of the canvas
canvas.grid(row=0,column=3,columnspan=2)
t.insert(str(1.0),est.info())
root.mainloop()
```

I worked on the GUI code above with ZhongYang. Our main goal was to show our graphs to the user. One of the most important aspects of the code was to display different graphs with user input. The most important part of the code was the estimator class and the usage of estimator button.

My main concern was to make the GUI as visually representable as possible. I especially worked on the canvasses and was involved in figuring out the sizing of the figures.

4. Results.

Results that I was able to observe was to be able to display the linear regression models via clicking the estimator button. The graphs showed the relationship between the certain features that we picked.

5. Summary and conclusions

The main lesson that I learned from this project is the fact that it was tough to work with multiple data-sets to build a model. It was very interesting to use the country codes to merge all of the data sets together while also selecting the features that the team liked. Following results and suggestions were achieved by the group:

- In the case of indoor pollution, mortality is expressed as a function of exponential decay until the consumption rate of clean fuel reaches very high values. Value> 94%.
- In the case of outdoor pollution, mortality has something to do with GDP / fuel.

According to the results that the team found, PollutionSolvers suggest the following improvements to the various countries.

- For countries with deaths due to the high indoor pollution, clean Fuel should be used more. Some examples of activities that create indoor pollution are cooking, warming up.

- For countries with deaths due to high outdoor pollution, energy density should be changed for activities that involve high energy consumption such as industrial production and transportation.

One thing that the team found out to be problematic was the fact that graphs were not enough for display as referenced by the professor. The team took note on that, and will try to do a better job in their next endeavors.

6. Percentage of Code

I’m not sure about this one as ZhongYang and I created this on our own. I did not personally copy and paste any code from the internet.

7. References