

# NLP (CSCE-689) – REPORT (Programming Assignment #1)

Name – Shiva Kumar Pentyala  
UIN - 127003995

---

## 1. Compile and Execution

I have used python 3.6. Follow the below Steps –

1. Unzip the file.
2. cd into this folder
3. python SpamLord.py data\_dev/dev/ data\_dev/devGOLD
4. Output will get printed on the terminal
5. Output - 3 categories
  - a. True Positive
  - b. False Positive
  - c. False Negative
  - d. Summary

## 2. Results and Analysis

True Positives (59):

dabo p 650-725-3897  
levoy p 650-725-4089  
latombe e liliana@cs.stanford.edu  
kosecka e kosecka@cs.gmu.edu  
ashishg p 650-814-1478  
bgirod p 650-723-4539  
eroberts e eroberts@cs.stanford.edu  
engler e engler@stanford.edu  
hanrahan e hanrahan@cs.stanford.edu  
hanrahan p 650-723-0033  
horowitz p 650-725-3707  
cheriton p 650-725-3726  
cheriton p 650-723-1131  
kunle e darlene@csl.stanford.edu  
cheriton e cheriton@cs.stanford.edu  
ashishg e rozm@stanford.edu  
kunle p 650-725-6949  
hager p 410-516-8000  
lam p 650-725-6949  
hager p 410-516-5553  
dlwh e dlwh@stanford.edu  
bgirod p 650-724-3648  
ashishg p 650-723-1614  
cheriton e uma@cs.stanford.edu  
bgirod p 650-724-6354  
jks e jks@robotics.stanford.edu  
lam e lam@cs.stanford.edu  
eroberts p 650-723-3642  
lam p 650-725-3714  
kosecka p 703-993-1876  
ashishg e ashishg@stanford.edu  
levoy e melissa@graphics.stanford.edu  
kunle p 650-723-1430  
eroberts p 650-723-6092  
fedkiw e fedkiw@cs.stanford.edu  
hager e hager@cs.jhu.edu  
engler e engler@lcs.mit.edu  
ashishg p 650-723-4173

dabo p 650-725-4671  
levoy p 650-723-0033  
dabo e dabo@cs.stanford.edu  
levoy p 650-724-6865  
hager p 410-516-5521  
hanrahan p 650-723-8530  
latombe p 650-723-4137  
kosecka p 703-993-1710  
balaji e balaji@stanford.edu  
latombe p 650-723-0350  
latombe e asandra@cs.stanford.edu  
kunle e kunle@ogun.stanford.edu  
jurafrsky e jurafrsky@stanford.edu  
horowitz p 650-725-6949  
levoy e ada@graphics.stanford.edu  
latombe p 650-721-6625  
latombe e latombe@cs.stanford.edu  
jurafrsky p 650-723-5666  
kunle p 650-725-3713  
latombe p 650-725-1449  
levoy p 650-725-3724

False Positives (0):

False Negatives (0):

Summary: tp=59, fp=0, fn=0

Analysis -

True Positive - 59

False Positive - 0

False Negative – 0

### 3. Limitations for this program –

The assumptions that were made according to the training data may produce false positives and may fail to get correct Results in some scenario's like below:

- A. It assumes a certain ordering in the format of phone numbers (specific to North America) and thus the SpamLord is not scalable.
  - B. If a line contains both Server and Port, that is where the address of a router or a machine is specified instead of email id, I have marked the detection as farce and skipped it. It may be that a real email is there and thus the SpamLord will leave a true positive.
  - C. For the part where email id is passed as parameters to the obfuscate function, a pattern in the function prototype was assumed. If it's changed or more parameters are passed, the SpamLord may give false positive and fail to find out the correct one.
- 

**Thank you so much**