

In [1]:

```
# Akshat Tandon 201503001
# Ques 1: Census Dataset
```

In [3]:

```
# importing modules
import numpy as np
import pandas as pd
import math
```

In [4]:

```
# reading the dataset into pandas dataframe
df=pd.read_csv("/home/tandon/IIIT-H/3rd/SMAI/census/census-income.data")
pd.options.display.max_columns = None
df.head()
```

Out[4]:

id	stock_dividends	tax_filer_stat	region_prev_res	state_prev_res	det_hh_fam_stat	...
0		Nonfiler	Not in universe	Not in universe	Other Rel 18+ ever marr not in subfamily	c
0		Head of household	South	Arkansas	Householder	t
0		Nonfiler	Not in universe	Not in universe	Child 18+ never marr Not in a subfamily	c
0		Nonfiler	Not in universe	Not in universe	Child <18 never marr not in subfamily	c
0		Nonfiler	Not in universe	Not in universe	Child <18 never marr not in subfamily	c

In [5]:

```
print('Total number of attributes', len(df.columns))
print()
print('Attribute Names: \n', df.columns)
```

Total number of attributes 42

Attribute Names:

```
Index(['age', 'class_worker', 'det_ind_code', 'det_occ_code', 'education',
       'wage_per_hour', 'hs_college', 'marital_stat', 'major_ind_code',
       'major_occ_code', 'race', 'hisp_origin', 'sex', 'union_member',
       'unemp_reason', 'full_or_part_emp', 'capital_gains', 'capital_losses',
       'stock_dividends', 'tax_filer_stat', 'region_prev_res', 'state_prev_res',
       'det_hh_fam_stat', 'det_hh_summ', 'income_value', 'mig_chg_msa',
       'mig_chg_reg', 'mig_move_reg', 'mig_same', 'mig_prev_sunbelt',
       'num_emp', 'fam_under_18', 'country_father', 'country_mother',
       'country_self', 'citizenship', 'own_or_self', 'vet_question',
       'vet_benefits', 'weeks_worked', 'year', 'income_50k'],
      dtype='object')
```

In [6]:

```
# deleting continuous attributes
del df['age']
del df['wage_per_hour']
del df['capital_gains']
del df['capital_losses']
del df['stock_dividends']
del df['num_emp']
del df['weeks_worked']
# deleting income_vale since it gives the exact income of the person
# which is not required since we have +50K and -50 class labels
del df['income_value']
print('Total number of attributes', len(df.columns))
df.head()
```

Total number of attributes 34

Out[6]:

	class_worker	det_ind_code	det_occ_code	education	hs_college	marital_stat	m
0	Not in universe	0	0	High school graduate	Not in universe	Widowed	Not in universe
1	Self-employed-not incorporated	4	34	Some college but no degree	Not in universe	Divorced	Children
2	Not in universe	0	0	10th grade	High school	Never married	Not in universe
3	Not in universe	0	0	Children	Not in universe	Never married	Not in universe
4	Not in universe	0	0	Children	Not in universe	Never married	Not in universe



In [7]:

```
# Trimming the white spaces around values
for attribute in df.columns:
    if df[attribute].dtype == 'object':
        df[attribute] = df[attribute].map(str.strip)
        print(attribute)
```

```
class_worker
education
hs_college
marital_stat
major_ind_code
major_occ_code
race
hisp_origin
sex
union_member
unemp_reason
full_or_part_emp
tax_filer_stat
region_prev_res
state_prev_res
det_hh_fam_stat
det_hh_summ
mig_chg_msa
    mig_chg_reg
mig_move_reg
mig_same
mig_prev_sunbelt
fam_under_18
country_father
country_mother
country_self
citizenship
vet_question
income_50k
```

In [8]:

```
# Calculate total number of likelihood values required for making the naive bayes classifier
counter = 0
for i in range(len(df.columns)):
    categories = df[df.columns[i]].unique()
    num_categories = len(categories)
    # print('For ',df.columns[i], ' num uniuue: ', num_categories)
    counter += num_categories

print('Total number of conditional values required: ', counter)
```

Total number of conditional values required: 505

In [9]:

```
# Playing around
print(len(df))

jemp = df.loc[df['country_father'] == 'United-States']
temp = jemp.loc[jemp['income_50k'] == '- 50000.']
kemp = jemp.loc[jemp['income_50k'] != '- 50000.']

print('Only USA' ,len(jemp))
print('USA and -50k', len(temp))
print('USA and +50k', len(kemp))

both = df.loc[(df['country_father'] == 'United-States') & (df['income_50k'] ==
'- 50000.')]
both1 = df.loc[(df['country_father'] == 'United-States') & (df['income_50k'] !=
'- 50000.')]
print('Both: USA and -50K', len(both))
print('Both: USA and +50K', len(both1))
```

```
199523
Only USA 159163
USA and -50k 148861
USA and +50k 10302
Both: USA and -50K 148861
Both: USA and +50K 10302
```

In [10]:

```
# Calculating priors
total_instances = len(df)
above50k = len(df.loc[df['income_50k'] != '- 50000.'])
below50k = total_instances - above50k
print('Total instaces', total_instances)
print('Num instances above 50k', above50k)
print('Num instance below 50k', below50k)

prior = {}
prior['+50k'] = above50k/total_instances
prior['-50k'] = below50k/total_instances

print('Prior probability +50k', prior['+50k'])
print('Prior probability -50k', prior['-50k'])
```

```
Total instaces 199523
Num instances above 50k 12382
Num instance below 50k 187141
Prior probability +50k 0.06205800834991455
Prior probability -50k 0.9379419916500854
```

In [11]:

```
# Calculating likelihood values
likelihood = {}

for i in range(2):
    if i == 0:
        category = '-50k'
        likelihood[category] = {}
        subframe = df.loc[df['income_50k'] == '- 50000.']
        total_category = len(subframe)
    else:
        category = '+50k'
        likelihood[category] = {}
        subframe = df.loc[df['income_50k'] != '- 50000.']
        total_category = len(subframe)
    for attribute in df.columns:
        print('\n')
        print('Attribute: ', attribute)
        likelihood[category][attribute] = {}
        for value in df[attribute].unique():
            num = len(subframe.loc[subframe[attribute] == value])
            likelihood[category][attribute][value] = num/total_category
            print(value, likelihood[category][attribute][value])
```


Attribute: class_worker
Not in universe 0.5308350388209959
Self-employed-not incorporated 0.039301916736578305
Private 0.3457606831212829
Local government 0.037068306784723816
Federal government 0.012439818105065165
Self-employed-incorporated 0.011387135902875371
State government 0.019995618277127942
Never worked 0.0023351376769387786
Without pay 0.0008763445744118071

Attribute: det_ind_code
0 0.5331701764979347
4 0.029127769970236347
40 0.008218402167349752
34 0.011964240866512417
43 0.03949962862226877
37 0.018713162802378953
24 0.007069535804553786
39 0.015132974602038036
12 0.005712270427111109
35 0.014892514200522601
45 0.018371174675779225
3 0.002228266387376363
19 0.006695486291085332
29 0.019947526196824855
32 0.01626581027139964
48 0.003136672348656895
33 0.08699857326828433
23 0.002313763419026296
44 0.01317188643856771
36 0.005022950609433529
31 0.004926766448827355
30 0.004867987239568026
41 0.018590260819382178
5 0.0027626228351884407
11 0.007694732848493917
9 0.004638213967008833
42 0.021817773764167124
6 0.002805371351013407
18 0.0023832297572418657
50 0.0075183952207159305
2 0.0110344606473194
1 0.004167980292934205
26 0.0004221415937715412
47 0.007106940755900631
16 0.0022336099518544842
14 0.0012076455720552953
22 0.004878674368524268
17 0.0007748168493275124
7 0.0020145238082515324
8 0.0024687267888917983
25 0.004221415937715412
46 0.0008603138809774448
27 0.0030244574946163586
15 0.0018007812291267013
13 0.0036710287964689726
49 0.0024526960954574357
38 0.008175653651524787

21 0.0028427763023602523
28 0.0007213812045463047
20 9.084059612805318e-05
51 0.0001496198053873817
10 2.13742579124831e-05

Attribute: det_occ_code

0 0.5331701764979347
34 0.01991546480995613
10 0.01766048060018916
3 0.013658150806076701
40 0.0031954515579162234
26 0.041129415788095605
37 0.011477976499003424
31 0.014272660721060591
12 0.015405496390422196
36 0.02135288365457062
41 0.008362678408259014
22 0.002051928759598378
2 0.0317140551776468
35 0.01530931222981602
25 0.0038473664242469584
23 0.017815443970054665
42 0.010072619041257662
8 0.00951154477105498
19 0.02818730262208709
29 0.027129276855419177
27 0.004157293163977963
16 0.015181066682341123
33 0.015875730064496822
13 0.006433651631657413
18 0.004200041679802929
9 0.0028801812537070978
17 0.0068931981767758
39 0.004932110013305476
32 0.012616155732843151
11 0.0011595534917522082
30 0.009992465574085849
38 0.01516503598890676
20 0.00035801882003409194
7 0.0012343633944458992
21 0.0024152911441105906
44 0.008378709101693376
24 0.009639790318529878
43 0.006764952629300902
28 0.007491677398325327
4 0.003660341667512731
1 0.0021962050005076388
6 0.0015336030052206626
45 0.0008122218006743579
14 0.004333630791755949
5 0.002874837689228977
15 0.003430568394953538
46 0.0001496198053873817

Attribute: education

High school graduate 0.24862536803800342
Some college but no degree 0.13910901405891815
10th grade 0.04005001576351521

Children 0.2534025146814434
Bachelors degree(BA AB BS) 0.08522985342602636
Masters degree(MA MS MEng MEd MSW MBA) 0.02406207084497785
Less than 1st grade 0.004371035743102794
Associates degree-academic program 0.021112423253055183
7th and 8th grade 0.04240118413388835
12th grade no diploma 0.011178736888228662
Associates degree-occup /vocational 0.026423926344307232
Prof school degree (MD DDS DVM LLB JD) 0.004403097129971519
5th or 6th grade 0.017393302376283122
11th grade 0.03636829983808999
Doctorate degree(PhD EdD) 0.00323820007374119
9th grade 0.03308735124852384
1st 2nd 3rd or 4th grade 0.009543606157923705

Attribute: hs_college

Not in universe 0.9329382657995843
High school 0.03680112856081778
College or university 0.03026060563959795

Attribute: marital_stat

Widowed 0.054146338856797815
Divorced 0.062220464783238304
Never married 0.45616941236821434
Married-civilian spouse present 0.39874746848632847
Separated 0.0176444499067548
Married-spouse absent 0.0075985486878877425
Married-A F spouse present 0.003473316910778504

Attribute: major_ind_code

Not in universe or children 0.5331701764979347
Construction 0.029127769970236347
Entertainment 0.008218402167349752
Finance insurance and real estate 0.026856755067035016
Education 0.03949962862226877
Business and repair services 0.02688881645390374
Manufacturing-nondurable goods 0.0322804730123276
Personal services except private HH 0.015132974602038036
Manufacturing-durable goods 0.04018894843994635
Other professional services 0.018371174675779225
Mining 0.002228266387376363
Transportation 0.019947526196824855
Wholesale trade 0.01626581027139964
Public administration 0.02021470442073089
Retail trade 0.08699857326828433
Social services 0.01317188643856771
Private household services 0.005022950609433529
Utilities and sanitary services 0.004926766448827355
Communications 0.004867987239568026
Hospital services 0.018590260819382178
Medical except hospital 0.021817773764167124
Agriculture 0.015202440940253605
Forestry and fisheries 0.0008603138809774448
Armed Forces 0.0001496198053873817

Attribute: major_occ_code

Not in universe 0.5331701764979347

Precision production craft & repair 0.05110050710426897
Professional specialty 0.05592040226353391
Executive admin and managerial 0.04756841098423114
Handlers equip cleaners etc 0.0216307490074329
Adm support including clerical 0.07689923640463607
Machine operators assmblrs & inspctrs 0.032830860153574046
Other service 0.06401055888340877
Sales 0.054819627981041036
Private household services 0.004157293163977963
Technicians and related support 0.0141978508183669
Transportation and material moving 0.020097146002212237
Farming forestry and fishing 0.015955883531668633
Protective services 0.007491677398325327
Armed Forces 0.0001496198053873817

Attribute: race

White 0.8340930100833062
Asian or Pacific Islander 0.02888196600424279
Amer Indian Aleut or Eskimo 0.011766528980821946
Black 0.10620334400265041
Other 0.019055150928978686

Attribute: hisp_origin

All other 0.8557184155262609
Do not know 0.001592382214479991
Central or South American 0.020385698484030757
Mexican (Mexicano) 0.03821717314751978
Mexican-American 0.04225156432850097
Other Spanish 0.012845929005402344
Puerto Rican 0.017339866731501917
Cuban 0.005690896169198626
Chicano 0.001592382214479991
NA 0.004365692178624673

Attribute: sex

Female 0.5414152964876751
Male 0.45858470351232494

Attribute: union_member

Not in universe 0.9100678098332273
No 0.07594273836305246
Yes 0.01398945180372019

Attribute: unemp_reason

Not in universe 0.9684729695790875
Job loser - on layoff 0.00509776051212722
Other job loser 0.010361171523076184
New entrant 0.0023351376769387786
Re-entrant 0.010665754698329068
Job leaver 0.003067206010441325

Attribute: full_or_part_emp

Not in labor force 0.14078154974056994
Children or Armed Forces 0.6299795341480487
Full-time schedules 0.18900187559113182

Unemployed full-time 0.011905461657253088
Unemployed part-time 0.00440844069444964
PT for non-econ reasons usually FT 0.01565130035641575
PT for econ reasons usually PT 0.005632116959939297
PT for econ reasons usually FT 0.002639720852191663

Attribute: tax_filer_stat
Nonfiler 0.4010826061632673
Head of household 0.03728739292832677
Joint both under 65 0.312758828904409
Single 0.18769804585847036
Joint both 65+ 0.04211797521654795
Joint one under 65 & one 65+ 0.019055150928978686

Attribute: region_prev_res
Not in universe 0.9190182803340796
South 0.025141470869558247
Northeast 0.013668837935032943
Midwest 0.018520794481166607
West 0.020914711367364714
Abroad 0.002735905012797837

Attribute: state_prev_res
Not in universe 0.9190182803340796
Arkansas 0.0010633693311460342
Utah 0.00551455854142064
Michigan 0.002254984209766967
Minnesota 0.002971021849835151
Alaska 0.001426731715658247
Kansas 0.0007748168493275124
Indiana 0.0027626228351884407
? 0.003542783248994074
Massachusetts 0.0007480990269369085
New Mexico 0.002313763419026296
Nevada 0.0009084059612805318
Tennessee 0.0010580257666679136
Colorado 0.0012183327010115367
Abroad 0.0034626297818222624
Kentucky 0.0012503940878802614
California 0.008800850695464917
Arizona 0.0012771119102708652
North Carolina 0.004109201083674876
Connecticut 0.0005610742702026814
Florida 0.004376379307580915
Vermont 0.0009939029929304643
Maryland 0.0007106940755900631
Oklahoma 0.00323820007374119
Oregon 0.0012183327010115367
Ohio 0.0011007742824928797
South Carolina 0.00048626436750899056
Texas 0.0010740564601022758
Montana 0.0011702406207084498
Wyoming 0.0012343633944458992
Georgia 0.001186271314142812
Pennsylvania 0.0010313079442773095
Iowa 0.0009885594284523434
New Hampshire 0.0012450505234021406
Missouri 0.0009084059612805318

Alabama 0.0011221485404053627
North Dakota 0.0025702545139760927
New Jersey 0.0003633623845122127
Louisiana 0.0010099336863648265
West Virginia 0.0011702406207084498
Delaware 0.0003740495134684543
Illinois 0.0009297802191930149
Maine 0.0008389396230649617
Wisconsin 0.0005343564478120775
New York 0.0010045901218867057
Idaho 0.00015496336986550247
District of Columbia 0.0005610742702026814
South Dakota 0.0007106940755900631
Nebraska 0.0009351237836711356
Virginia 0.0006572584308088554
Mississippi 0.0010633693311460342

Attribute: det_hh_fam_stat

Other Rel 18+ ever marr not in subfamily 0.010355827958598062
Householder 0.24258714017772695
Child 18+ never marr Not in a subfamily 0.06381819056219643
Child <18 never marr not in subfamily 0.2689095387969499
Spouse of householder 0.2106059067761741
Secondary individual 0.03153771754986882
Other Rel 18+ never marr not in subfamily 0.009116120999674043
Nonfamily householder 0.10907818169187938
Grandchild <18 never marr not in subfamily 0.005696239733676747
Grandchild <18 never marr child of subfamily RP 0.009981778445129609
Child 18+ ever marr Not in a subfamily 0.005290128833339567
Child 18+ never marr RP of subfamily 0.003136672348656895
Child 18+ spouse of subfamily RP 0.0006626019952869762
Other Rel <18 never marr child of subfamily RP 0.0035053782976472284
Child under 18 of RP of unrel subfamily 0.003911489197984407
Grandchild 18+ never marr not in subfamily 0.001993149550339049
Child 18+ ever marr RP of subfamily 0.0035214089910815907
Other Rel 18+ ever marr RP of subfamily 0.0034038505725629336
RP of unrelated subfamily 0.003612249587209644
Other Rel 18+ spouse of subfamily RP 0.003280948589566156
Other Rel <18 never marr not in subfamily 0.003120641655222533
Other Rel <18 spouse of subfamily RP 1.6030693434362327e-05
In group quarters 0.0010366515087554303
Grandchild 18+ spouse of subfamily RP 5.3435644781207754e-05
Other Rel 18+ never marr RP of subfamily 0.0005022950609433528
Child <18 never marr RP of subfamily 0.00042748515824966204
Child <18 ever marr not in subfamily 0.0001923683212123479
Other Rel <18 ever marr RP of subfamily 3.2061386868724654e-05
Grandchild 18+ ever marr not in subfamily 0.0001763376277779856
Child <18 spouse of subfamily RP 1.068712895624155e-05
Spouse of RP of unrelated subfamily 0.0002511475304716764
Other Rel <18 never married RP of subfamily 2.13742579124831e-05
Grandchild 18+ never marr RP of subfamily 3.2061386868724654e-05
Grandchild 18+ ever marr RP of subfamily 4.8092080303086974e-05
Child <18 ever marr RP of subfamily 4.8092080303086974e-05
Other Rel <18 ever marr not in subfamily 5.343564478120775e-06
Grandchild <18 never marr RP of subfamily 1.068712895624155e-05
Grandchild <18 ever marr not in subfamily 1.068712895624155e-05

Attribute: det_hh_summ

Other relative of householder 0.0513890595860875

Householder 0.3517347882078219
Child 18 or older 0.07643434629503956
Child under 18 never married 0.269443895244762
Spouse of householder 0.2106807166788678
Nonrelative of householder 0.03936603951031575
Group Quarters- Secondary individual 0.0007000069466338215
Child under 18 ever married 0.0002511475304716764

Attribute: mig_chg_msa
? 0.49776906183038455
MSA to MSA 0.05407152895410412
Nonmover 0.413148374754864
NonMSA to nonMSA 0.014753581524091461
Not in universe 0.008100843748831096
Not identifiable 0.0022068921294638804
Abroad to MSA 0.0023244505479825374
MSA to nonMSA 0.004093170390240514
Abroad to nonMSA 0.0003900802069028166
NonMSA to MSA 0.0031420159131350157

Attribute: mig_chg_reg
? 0.49776906183038455
Same county 0.05067836551049743
Nonmover 0.413148374754864
Different region 0.005893951619367215
Different county same state 0.014267317156582469
Not in universe 0.008100843748831096
Different division same region 0.0023511683703731413
Abroad 0.002735905012797837
Different state same division 0.005055011996302253

Attribute: mig_move_reg
? 0.49776906183038455
Same county 0.05067836551049743
Nonmover 0.413148374754864
Different state in South 0.004873330804046147
Different county same state 0.014267317156582469
Not in universe 0.008100843748831096
Different state in Northeast 0.0021588000491607933
Abroad 0.002735905012797837
Different state in Midwest 0.002810714915491528
Different state in West 0.0034572862173441418

Attribute: mig_same
Not in universe under 1 year old 0.5058699055792156
No 0.08098171966592035
Yes 0.413148374754864

Attribute: mig_prev_sunbelt
? 0.49776906183038455
Yes 0.029881212561651375
Not in universe 0.42124921850369507
No 0.05110050710426897

Attribute: fam_under_18

Not in universe 0.7045596635691804
Both parents present 0.20829748692162595
Mother only present 0.06824800551455854
Neither parent present 0.008832912082333642
Father only present 0.01006193191230142

Attribute: country_father

United-States 0.7954483517775367
Vietnam 0.0023832297572418657
Philippines 0.00571761399158923
? 0.032937731443136456
Columbia 0.003206138686872465
Germany 0.006658081339738486
Mexico 0.05281579130174574
Japan 0.0018649040028641507
Peru 0.0017420020198673728
Dominican-Republic 0.006818388274082109
South Korea 0.002661095110104146
Cuba 0.005696239733676747
El-Salvador 0.005209975366167756
Canada 0.0066794555976509695
Scotland 0.001212989136533416
Outlying-U S (Guam USVI etc) 0.0008068782361962371
Italy 0.011050491340753763
Guatemala 0.0023671990638075035
Ecuador 0.0019664317279484453
Puerto-Rico 0.013968077545807707
Cambodia 0.0010206208153210681
China 0.004157293163977963
Poland 0.005931356570714061
Nicaragua 0.001640474294783078
Taiwan 0.0009564980415836188
England 0.0038153050373782335
Ireland 0.0023671990638075035
Hungary 0.0014534495380488508
Yugoslavia 0.0010794000245803966
Trinidad&Tobago 0.000593135657071406
Jamaica 0.002356511934851262
Honduras 0.0010313079442773095
Portugal 0.0019664317279484453
Iran 0.0010847435890585175
France 0.000951154477105498
India 0.0026771258035385086
Hong Kong 0.0005397000122901983
Haiti 0.0018274990515173052
Greece 0.0016297871658268364
Holand-Netherlands 0.0002351168370373141
Thailand 0.0005290128833339568
Laos 0.0008122218006743579
Panama 0.00013358911195301938

Attribute: country_mother

United-States 0.8020209360856253
Vietnam 0.0024687267888917983
? 0.03019113930138238
Columbia 0.003206138686872465
Mexico 0.05153867939147488
El-Salvador 0.005813798152195403
Peru 0.0018488733094297882

Puerto-Rico 0.012931426037052276
 Cuba 0.005589368444114331
 Philippines 0.006155786278795133
 Dominican-Republic 0.0058351724101078865
 Germany 0.006738234806910297
 England 0.004296225840409103
 Guatemala 0.002356511934851262
 Scotland 0.001143522798317846
 Portugal 0.0017420020198673728
 Italy 0.00930314575640827
 Ecuador 0.0019397139055578414
 Yugoslavia 0.0008923752678461694
 China 0.003687059489903335
 Poland 0.005418374380814466
 Hungary 0.0014107010222238847
 Nicaragua 0.0015763515210456287
 Taiwan 0.0010794000245803966
 Ireland 0.002810714915491528
 Canada 0.007026787288728819
 South Korea 0.003067206010441325
 Trinidad&Tobago 0.0005183257543777152
 Jamaica 0.0023084198545481748
 Honduras 0.0011542099272740874
 Iran 0.0009137495257586526
 France 0.000999246557408585
 Cambodia 0.0008068782361962371
 India 0.0026717822390603875
 Hong Kong 0.0005343564478120775
 Haiti 0.0018435297449516674
 Japan 0.0022229228228982426
 Greece 0.0012771119102708652
 Holand-Netherlands 0.00022442970808107256
 Thailand 0.0006198534794620099
 Panama 0.0001709940632998648
 Laos 0.0008175653651524786
 Outlying-U S (Guam USVI etc) 0.0008282524941087201

Attribute: country_self

United-States 0.8861767330515493
 Vietnam 0.0020305545016858946
 ? 0.01654367562426192
 Columbia 0.0022710149032013293
 Mexico 0.030415569009463454
 Peru 0.0013839831998332809
 Cuba 0.004151949599499842
 Philippines 0.004098513954718635
 Dominican-Republic 0.0036282802806440065
 El-Salvador 0.003612249587209644
 Canada 0.003371789185694209
 Scotland 0.0003313009976434881
 Portugal 0.0008816881388899279
 Guatemala 0.0018221554870391844
 Ecuador 0.001330547555052073
 Germany 0.004162636728456084
 Outlying-U S (Guam USVI etc) 0.0006198534794620099
 Puerto-Rico 0.007320683335025462
 Italy 0.00206795945303274
 China 0.0023671990638075035
 Poland 0.0019129960831672376
 Nicaragua 0.0011542099272740874

Taiwan 0.0009084059612805318
England 0.0021320822267701894
Ireland 0.0006465713018526138
South Korea 0.0023618554993293825
Trinidad&Tobago 0.00034733169107785037
Jamaica 0.0016137564723924741
Honduras 0.0007587861558931501
Iran 0.00068931981767758
Hungary 0.00037939307794657503
France 0.0005664178346808022
Cambodia 0.0005022950609433528
India 0.0018007812291267013
Hong Kong 0.0004809208030308698
Japan 0.0015977257789581117
Haiti 0.0011755841851865706
Holand-Netherlands 9.618416060617395e-05
Greece 0.0007053505111119424
Thailand 0.0005557307057245607
Panama 0.0001496198053873817
Yugoslavia 0.0003206138686872465
Laos 0.0005557307057245607

Attribute: citizenship

Native- Born in the United States 0.8861927637449837
Foreign born- Not a citizen of U S 0.06898007384806108
Foreign born- U S citizen by naturalization 0.02818730262208709
Native- Born abroad of American Parent(s) 0.008699322970380622
Native- Born in Puerto Rico or U S Outlying 0.007940536814487472

Attribute: own_or_self

0 0.9095815454657183
2 0.07925574833948734
1 0.0111627061947943

Attribute: vet_question

Not in universe 0.9906327314698543
No 0.0073634318508504285
Yes 0.0020038366792952907

Attribute: vet_benefits

2 0.7372996831266264
0 0.25333304834322784
1 0.009367268530145719

Attribute: year

95 0.49776906183038455
94 0.5022309381696154

Attribute: income_50k

- 50000. 1.0
50000+. 0.0

Attribute: class_worker

Not in universe 0.07300920691326118

Self-employed-not incorporated 0.08803101276045873
Private 0.5913422710386044
Local government 0.06840575028266839
Federal government 0.04821515102568244
Self-employed-incorporated 0.09158455822968825
State government 0.039169762558552736
Never worked 0.00016152479405588758
Without pay 8.076239702794379e-05

Attribute: det_ind_code
0 0.07317073170731707
4 0.04304635761589404
40 0.009126150864157648
34 0.042481020836698434
43 0.07195929575189791
37 0.041996446454530774
24 0.014537231465029882
39 0.008480051687934097
12 0.022694233564852204
35 0.04789210143757067
45 0.08431594249717332
3 0.011791309966079794
19 0.007510902923598773
29 0.038442900985301245
32 0.044580843159424975
48 0.005249555806816347
33 0.06372153125504765
23 0.007430140526570829
44 0.006784041350347278
36 0.00040381198513971895
31 0.02067517363915361
30 0.021805847197544824
41 0.039169762558552736
5 0.0029074462930059764
11 0.02616701663705379
9 0.010095299628492973
42 0.04845743821676628
6 0.00234210951381037
18 0.00298820869003392
50 0.023986431917299305
2 0.010579874010660637
1 0.003795832660313358
26 0.003876595057341302
47 0.02535939266677435
16 0.009772250040381198
14 0.005572605394928122
22 0.0031497334840898078
17 0.0009691487643353255
7 0.0036343078662574705
8 0.0071070909384590535
25 0.023744144726215475
46 0.0020998223227265385
27 0.004845743821676627
15 0.009287675658213536
13 0.017121628169924085
49 0.012195121951219513
38 0.007995477305766436
21 0.0021805847197544823
28 0.0006460991762235503
20 0.0012114359554191568

51 0.0006460991762235503
10 0.0

Attribute: det_occ_code

0 0.07317073170731707
34 0.02406719431432725
10 0.03052818607656275
3 0.05160717170085608
40 0.001534485543530932
26 0.01534485543530932
37 0.006945566144403166
31 0.002261347116782426
12 0.036908415441770315
36 0.012033597157163624
41 0.0021805847197544823
22 0.0021805847197544823
2 0.22783072201582943
35 0.02447100629946697
25 0.003795832660313358
23 0.00468421902762074
42 0.002665159101922145
8 0.029962849297367145
19 0.011145210789856244
29 0.002261347116782426
27 0.00016152479405588758
16 0.04878048780487805
33 0.0285898885478921
13 0.005411080600872234
18 0.023986431917299305
9 0.016071717008560815
17 0.03884671297044096
39 0.007591665320626716
32 0.00298820869003392
11 0.03392020675173639
30 0.0021805847197544823
38 0.013325795509610725
20 0.00032304958811177516
7 0.04038119851397189
21 0.006541754159263447
44 0.001938297528670651
24 0.003472783072201583
43 0.00936843805524148
28 0.02091746083023744
4 0.05483766758197383
1 0.010741398804716524
6 0.012437409142303344
45 0.0016152479405588758
14 0.009772250040381198
5 0.025601679857858183
15 0.013971894685834275
46 0.0006460991762235503

Attribute: education

High school graduate 0.1517525440155064
Some college but no degree 0.14432240348893555
10th grade 0.005007268615732515
Children 0.0
Bachelors degree(BA AB BS) 0.31618478436439995
Masters degree(MA MS MEng MEd MSW MBA) 0.16459376514294943

Less than 1st grade 8.076239702794379e-05
Associates degree-academic program 0.03327410757551284
7th and 8th grade 0.005814892586011953
12th grade no diploma 0.002745921498950089
Associates degree-occup /vocational 0.033354869972540786
Prof school degree (MD DDS DVM LLB JD) 0.07825876272007753
5th or 6th grade 0.0017767727346147634
11th grade 0.005653367791956065
Doctorate degree(PhD EdD) 0.05306089484735907
9th grade 0.003068971087061864
1st 2nd 3rd or 4th grade 0.0010499111613632693

Attribute: hs_college
Not in universe 0.9975771280891617
High school 0.00040381198513971895
College or university 0.0020190599256985947

Attribute: marital_stat
Widowed 0.02665159101922145
Divorced 0.08609271523178808
Never married 0.09021159748021321
Married-civilian spouse present 0.7753190114682604
Separated 0.01276045873041512
Married-spouse absent 0.007753190114682604
Married-A F spouse present 0.0012114359554191568

Attribute: major_ind_code
Not in universe or children 0.07317073170731707
Construction 0.04304635761589404
Entertainment 0.009126150864157648
Finance insurance and real estate 0.0903731222742691
Education 0.07195929575189791
Business and repair services 0.04999192376029721
Manufacturing-nondurable goods 0.06913261185591989
Personal services except private HH 0.008480051687934097
Manufacturing-durable goods 0.12065902115974803
Other professional services 0.08431594249717332
Mining 0.011791309966079794
Transportation 0.038442900985301245
Wholesale trade 0.044580843159424975
Public administration 0.06679050234210951
Retail trade 0.06372153125504765
Social services 0.006784041350347278
Private household services 0.00040381198513971895
Utilities and sanitary services 0.02067517363915361
Communications 0.021805847197544824
Hospital services 0.039169762558552736
Medical except hospital 0.04845743821676628
Agriculture 0.014375706670973995
Forestry and fisheries 0.0020998223227265385
Armed Forces 0.0006460991762235503

Attribute: major_occ_code
Not in universe 0.07317073170731707
Precision production craft & repair 0.07712808916168631
Professional specialty 0.28064932967210465
Executive admin and managerial 0.290179292521402

Handlers equip cleaners etc 0.00638022936520756
Adm support including clerical 0.03602002907446293
Machine operators assmblrs & inspctrs 0.01897916330156679
Other service 0.009691487643353255
Sales 0.12308189307058634
Private household services 0.00016152479405588758
Technicians and related support 0.029155225327087708
Transportation and material moving 0.02091746083023744
Farming forestry and fishing 0.012921983524471006
Protective services 0.02091746083023744
Armed Forces 0.0006460991762235503

Attribute: race

White 0.9103537392989824
Asian or Pacific Islander 0.03472783072201583
Amer Indian Aleut or Eskimo 0.003957357454369246
Black 0.04361169439508965
Other 0.007349378129542885

Attribute: hisp_origin

All other 0.9503311258278145
Do not know 0.0006460991762235503
Central or South American 0.0006460991762235503
Mexican (Mexicano) 0.006622516556291391
Mexican-American 0.013891132288806332
Other Spanish 0.006541754159263447
Puerto Rican 0.005491842997900178
Cuban 0.004926506218704572
Chicano 0.00048457438216766274
NA 0.004603456630592796

Attribute: sex

Female 0.21507026328541431
Male 0.7849297367145857

Attribute: union_member

Not in universe 0.8195768050395735
No 0.14714908738491358
Yes 0.03327410757551284

Attribute: unemp_reason

Not in universe 0.9862703925052495
Job loser - on layoff 0.0017767727346147634
Other job loser 0.007995477305766436
New entrant 0.00016152479405588758
Re-entrant 0.0018575351316427072
Job leaver 0.001938297528670651

Attribute: full_or_part_emp

Not in labor force 0.03731222742691003
Children or Armed Forces 0.47439832014214184
Full-time schedules 0.43337102245194636
Unemployed full-time 0.006703278953319335
Unemployed part- time 0.0014537231465029882
PT for non-econ reasons usually FT 0.03173962203198191

PT for econ reasons usually PT 0.012518171539331288
PT for econ reasons usually FT 0.0025036343078662575

Attribute: tax_filer_stat

Nonfiler 0.0028266838959780326
Head of household 0.03618155386851882
Joint both under 65 0.7149895008883864
Single 0.185349701179131
Joint both 65+ 0.036343078662574704
Joint one under 65 & one 65+ 0.024309481505411082

Attribute: region_prev_res

Not in universe 0.9500888386367308
South 0.014860281053141657
Northeast 0.011872072363107737
Midwest 0.008803101276045873
West 0.012921983524471006
Abroad 0.0014537231465029882

Attribute: state_prev_res

Not in universe 0.9500888386367308
Arkansas 0.00048457438216766274
Utah 0.0025036343078662575
Michigan 0.001534485543530932
Minnesota 0.0016152479405588758
Alaska 0.0018575351316427072
Kansas 0.00032304958811177516
Indiana 0.0012921983524471006
? 0.0036343078662574705
Massachusetts 0.0008883863673073817
New Mexico 0.0024228719108383137
Nevada 0.00032304958811177516
Tennessee 0.00032304958811177516
Colorado 0.0008883863673073817
Abroad 0.0018575351316427072
Kentucky 0.0008076239702794379
California 0.005411080600872234
Arizona 0.00032304958811177516
North Carolina 0.003472783072201583
Connecticut 0.0009691487643353255
Florida 0.0024228719108383137
Vermont 0.00040381198513971895
Maryland 0.00024228719108383137
Oklahoma 0.0016152479405588758
Oregon 0.0006460991762235503
Ohio 0.00040381198513971895
South Carolina 0.00032304958811177516
Texas 0.0006460991762235503
Montana 0.0005653367791956065
Wyoming 0.0008076239702794379
Georgia 0.00040381198513971895
Pennsylvania 0.00048457438216766274
Iowa 0.00032304958811177516
New Hampshire 0.0007268615732514941
Missouri 0.00040381198513971895
Alabama 0.00048457438216766274
North Dakota 0.0014537231465029882
New Jersey 0.0005653367791956065

Louisiana 0.00024228719108383137
West Virginia 0.0009691487643353255
Delaware 0.00024228719108383137
Illinois 0.00048457438216766274
Maine 0.0008076239702794379
Wisconsin 0.00040381198513971895
New York 0.0005653367791956065
Idaho 0.00016152479405588758
District of Columbia 0.0008883863673073817
South Dakota 0.00040381198513971895
Nebraska 0.00024228719108383137
Virginia 0.00024228719108383137
Mississippi 0.00040381198513971895

Attribute: det_hh_fam_stat

Other Rel 18+ ever marr not in subfamily 0.0014537231465029882
Householder 0.6339848166693588
Child 18+ never marr Not in a subfamily 0.00702632854143111
Child <18 never marr not in subfamily 0.00016152479405588758
Spouse of householder 0.18429979001776772
Secondary individual 0.017767727346147635
Other Rel 18+ never marr not in subfamily 0.0017767727346147634
Nonfamily householder 0.14537231465029882
Grandchild <18 never marr not in subfamily 0.0
Grandchild <18 never marr child of subfamily RP 0.0
Child 18+ ever marr Not in a subfamily 0.0018575351316427072
Child 18+ never marr RP of subfamily 0.00016152479405588758
Child 18+ spouse of subfamily RP 0.00016152479405588758
Other Rel <18 never marr child of subfamily RP 0.0
Child under 18 of RP of unrel subfamily 0.0
Grandchild 18+ never marr not in subfamily 0.00016152479405588758
Child 18+ ever marr RP of subfamily 0.0009691487643353255
Other Rel 18+ ever marr RP of subfamily 0.001534485543530932
RP of unrelated subfamily 0.0007268615732514941
Other Rel 18+ spouse of subfamily RP 0.001938297528670651
Other Rel <18 never marr not in subfamily 0.0
Other Rel <18 spouse of subfamily RP 0.0
In group quarters 0.00016152479405588758
Grandchild 18+ spouse of subfamily RP 0.0
Other Rel 18+ never marr RP of subfamily 0.0
Child <18 never marr RP of subfamily 0.0
Child <18 ever marr not in subfamily 0.0
Other Rel <18 ever marr RP of subfamily 0.0
Grandchild 18+ ever marr not in subfamily 8.076239702794379e-05
Child <18 spouse of subfamily RP 0.0
Spouse of RP of unrelated subfamily 0.00040381198513971895
Other Rel <18 never married RP of subfamily 0.0
Grandchild 18+ never marr RP of subfamily 0.0
Grandchild 18+ ever marr RP of subfamily 0.0
Child <18 ever marr RP of subfamily 0.0
Other Rel <18 ever marr not in subfamily 0.0
Grandchild <18 never marr RP of subfamily 0.0
Grandchild <18 ever marr not in subfamily 0.0

Attribute: det_hh_summ

Other relative of householder 0.006945566144403166
Householder 0.7794378937166855
Child 18 or older 0.010176062025520918
Child under 18 never married 0.00016152479405588758

Spouse of householder 0.18429979001776772
Nonrelative of householder 0.018898400904538846
Group Quarters- Secondary individual 8.076239702794379e-05
Child under 18 ever married 0.0

Attribute: mig_chg_msa
? 0.5284283637538362
MSA to MSA 0.038927475367468906
Nonmover 0.42166047488289454
NonMSA to nonMSA 0.0040381198513971895
Not in universe 0.0
Not identifiable 0.0013729607494750444
Abroad to MSA 0.0014537231465029882
MSA to nonMSA 0.001938297528670651
Abroad to nonMSA 0.0
NonMSA to MSA 0.0021805847197544823

Attribute: mig_chg_reg
? 0.5284283637538362
Same county 0.026490066225165563
Nonmover 0.42166047488289454
Different region 0.006057179777095785
Different county same state 0.010256824422548862
Not in universe 0.0
Different division same region 0.0020190599256985947
Abroad 0.0014537231465029882
Different state same division 0.0036343078662574705

Attribute: mig_move_reg
? 0.5284283637538362
Same county 0.026490066225165563
Nonmover 0.42166047488289454
Different state in South 0.004926506218704572
Different county same state 0.010256824422548862
Not in universe 0.0
Different state in Northeast 0.0021805847197544823
Abroad 0.0014537231465029882
Different state in Midwest 0.0020190599256985947
Different state in West 0.0025843967048942013

Attribute: mig_same
Not in universe under 1 year old 0.5284283637538362
No 0.049911161363269264
Yes 0.42166047488289454

Attribute: mig_prev_sunbelt
? 0.5284283637538362
Yes 0.015667905023421094
Not in universe 0.42166047488289454
No 0.03424325633984817

Attribute: fam_under_18
Not in universe 0.9998384752059442
Both parents present 0.00016152479405588758
Mother only present 0.0

Neither parent present 0.0
Father only present 0.0

Attribute: country_father

United-States 0.8320142141818769
Vietnam 0.0008883863673073817
Philippines 0.006784041350347278
? 0.04433855596834114
Columbia 0.001130673558391213
Germany 0.008883863673073818
Mexico 0.01001453723146503
Japan 0.003472783072201583
Peru 0.0007268615732514941
Dominican-Republic 0.001130673558391213
South Korea 0.0025843967048942013
Cuba 0.004764981424648684
El-Salvador 0.0005653367791956065
Canada 0.010499111613632693
Scotland 0.0016152479405588758
Outlying-U S (Guam USVI etc) 0.0006460991762235503
Italy 0.011629785172023906
Guatemala 0.00016152479405588758
Ecuador 0.0008883863673073817
Puerto-Rico 0.00533031820384429
Cambodia 0.00040381198513971895
China 0.0062994669681796156
Poland 0.008237764496850267
Nicaragua 0.0006460991762235503
Taiwan 0.0016152479405588758
England 0.00638022936520756
Ireland 0.005249555806816347
Hungary 0.002745921498950089
Yugoslavia 0.0012114359554191568
Trinidad&Tobago 0.00016152479405588758
Jamaica 0.0017767727346147634
Honduras 8.076239702794379e-05
Portugal 0.0016152479405588758
Iran 0.0024228719108383137
France 0.0010499111613632693
India 0.00638022936520756
Hong Kong 0.00040381198513971895
Haiti 0.0007268615732514941
Greece 0.0031497334840898078
Holand-Netherlands 0.0005653367791956065
Thailand 0.0006460991762235503
Laos 0.00016152479405588758
Panama 0.0

Attribute: country_mother

United-States 0.83895978032628
Vietnam 0.0008883863673073817
? 0.037877564206105635
Columbia 0.0009691487643353255
Mexico 0.010983685995800355
El-Salvador 0.0016152479405588758
Peru 0.0007268615732514941
Puerto-Rico 0.004280407042481021
Cuba 0.005007268615732515
Philippines 0.00638022936520756

Dominican-Republic 0.0008883863673073817
Germany 0.009772250040381198
England 0.007995477305766436
Guatemala 0.00024228719108383137
Scotland 0.0021805847197544823
Portugal 0.0012921983524471006
Italy 0.00831852689387821
Ecuador 0.0009691487643353255
Yugoslavia 0.0008076239702794379
China 0.005653367791956065
Poland 0.007753190114682604
Hungary 0.002665159101922145
Nicaragua 0.00048457438216766274
Taiwan 0.0016152479405588758
Ireland 0.005895654983039897
Canada 0.010983685995800355
South Korea 0.0028266838959780326
Trinidad&Tobago 0.00016152479405588758
Jamaica 0.0016960103375868196
Honduras 0.00016152479405588758
Iran 0.0021805847197544823
France 0.0020190599256985947
Cambodia 0.00048457438216766274
India 0.006541754159263447
Hong Kong 0.0005653367791956065
Haiti 0.0006460991762235503
Japan 0.004280407042481021
Greece 0.0017767727346147634
Holand-Netherlands 0.0005653367791956065
Thailand 0.0005653367791956065
Panama 0.0
Laos 0.00016152479405588758
Outlying-U S (Guam USVI etc) 0.00016152479405588758

Attribute: country_self
United-States 0.9004199644645453
Vietnam 0.0008883863673073817
? 0.023986431917299305
Columbia 0.0007268615732514941
Mexico 0.006057179777095785
Peru 0.0007268615732514941
Cuba 0.004845743821676627
Philippines 0.0062994669681796156
Dominican-Republic 0.0008883863673073817
El-Salvador 0.0010499111613632693
Canada 0.005572605394928122
Scotland 0.0010499111613632693
Portugal 0.0007268615732514941
Guatemala 0.00024228719108383137
Ecuador 0.0007268615732514941
Germany 0.005814892586011953
Outlying-U S (Guam USVI etc) 0.00024228719108383137
Puerto-Rico 0.0024228719108383137
Italy 0.0025843967048942013
China 0.0028266838959780326
Poland 0.0018575351316427072
Nicaragua 0.00016152479405588758
Taiwan 0.0025036343078662575
England 0.00468421902762074
Ireland 0.001130673558391213

South Korea 0.00234210951381037
Trinidad&Tobago 8.076239702794379e-05
Jamaica 0.0014537231465029882
Honduras 0.00016152479405588758
Iran 0.002261347116782426
Hungary 0.0006460991762235503
France 0.0012114359554191568
Cambodia 8.076239702794379e-05
India 0.0057341301889840095
Hong Kong 0.0008076239702794379
Japan 0.0032304958811177516
Haiti 0.0006460991762235503
Holand-Netherlands 0.00040381198513971895
Greece 0.0012114359554191568
Thailand 0.0007268615732514941
Panama 0.0
Yugoslavia 0.00048457438216766274
Laos 8.076239702794379e-05

Attribute: citizenship

Native- Born in the United States 0.9004199644645453
Foreign born- Not a citizen of U S 0.039735099337748346
Foreign born- U S citizen by naturalization 0.046842190276207396
Native- Born abroad of American Parent(s) 0.010337586819576805
Native- Born in Puerto Rico or U S Outlying 0.002665159101922145

Attribute: own_or_self

0 0.8441285737360685
2 0.10668712647391375
1 0.04918429979001777

Attribute: vet_question

Not in universe 0.981343886286545
No 0.017363915361007915
Yes 0.0012921983524471006

Attribute: vet_benefits

2 0.981343886286545
0 0.0
1 0.018656113713455016

Attribute: year

95 0.5284283637538362
94 0.4715716362461638

Attribute: income_50k

- 50000. 0.0
50000+. 1.0

In [12]:

```
tf=pd.read_csv("/home/tandon/IIIT-H/3rd/SMAI/census/census-income.test")
tf.head()
```

Out[12]:

	age	class_worker	det_ind_code	det_occ_code	education	wage_per_hour	hs_c
0	38	Private	6	36	1st 2nd 3rd or 4th grade	0	Not in universe
1	44	Self- employed-not incorporated	37	12	Associates degree- occup /vocational	0	Not in universe
2	2	Not in universe	0	0	Children	0	Not in universe
3	35	Private	29	3	High school graduate	0	Not in universe
4	49	Private	4	34	High school graduate	0	Not in universe

In [13]:

```
print('Total number of attributes', len(tf.columns))
print()
print('Attribute Names: \n', tf.columns)
```

Total number of attributes 42

Attribute Names:

```
Index(['age', 'class_worker', 'det_ind_code', 'det_occ_code', 'education',
       'wage_per_hour', 'hs_college', 'marital_stat', 'major_ind_code',
       'major_occ_code', 'race', 'hisp_origin', 'sex', 'union_member',
       'unemp_reason', 'full_or_part_emp', 'capital_gains', 'capital_losses',
       'stock_dividends', 'tax_filer_stat', 'region_prev_res',
       'state_prev_res', 'det_hh_fam_stat', 'det_hh_summ', 'income_value',
       'mig_chg_msa', 'mig_chg_reg', 'mig_move_reg', 'mig_same',
       'mig_prev_sunbelt', 'num_emp', 'fam_under_18', 'country_father',
       'country_mother', 'country_self', 'citizenship', 'own_or_self',
       'vet_question', 'vet_benefits', 'weeks_worked', 'year', 'income_50k'],
      dtype='object')
```

In [14]:

```
# deleting continuous attributes
del tf['age']
del tf['wage_per_hour']
del tf['capital_gains']
del tf['capital_losses']
del tf['stock_dividends']
del tf['num_emp']
del tf['weeks_worked']
# deleting income_vale since it gives the exact income of the person
# which is not required since we have +50K and -50K class labels
del tf['income_value']
# removing the class labels from test data and storing it into another list
test_categories = tf['income_50k']
del tf['income_50k']
print('Total number of attributes', len(tf.columns))
tf.head()
# print('Income Values \n')
print(test_categories[6])
```

Total number of attributes 33
- 50000.

In [15]:

```
# Trimming the white spaces around values
for attribute in tf.columns:
    if tf[attribute].dtype == 'object':
        tf[attribute] = tf[attribute].map(str.strip)
        print(attribute)
test_categories = test_categories.map(str.strip)
```

```
class_worker
education
hs_college
marital_stat
major_ind_code
major_occ_code
race
hisp_origin
sex
union_member
unemp_reason
full_or_part_emp
tax_filer_stat
region_prev_res
state_prev_res
det_hh_fam_stat
det_hh_summ
mig_chg_msa
    mig_chg_reg
mig_move_reg
mig_same
mig_prev_sunbelt
fam_under_18
country_father
country_mother
country_self
citizenship
vet_question
```

In [18]:

```

correct = 0
label = ''
print('Validating')
for index, row in tf.iterrows():
#     print('\n')
#     print('Index: ', index)

    categories = ['-50k', '+50k']
    posterior = {}
    for category in categories:
#         print('Category', category)
        sum1 = 0
        posterior[category] = 0
        for attribute in tf.columns:
            num = likelihood[category][attribute][row[attribute]]
#             print('Attribute', attribute, num)
            if num == 0.0:
#                 print('0 reached')
                sum1 = -1 * math.inf
                break
            sum1 += math.log(num)
        posterior[category] = sum1 + math.log(prior[category])
#         print('Posterior +50k', posterior['+50k'])
#         print('Posterior -50k', posterior['-50k'])
    if posterior['-50k'] >= posterior['+50k']:
        label = '- 50000.'
    else:
        label = '50000+.'

# validating
#     print('Label assigned', label)
    if label == test_categories[index]:
        correct += 1

percent_correct = correct/len(tf) * 100
print('Percent correct', percent_correct)

```

Validating

Percent correct 77.88135763116216

In []: