Abhineet Pandey, 2017csb1062
Part 1
1)
A)
Deciding alpha: As the dataset was large, very small alpha would take too much time. Varied the alpha values multiplying by 10 each time, and observed the effect in some iterations.

B) Results:

Lambda	Error after 90 iterations
2	0.4019315537
5	0.3987769015
10	0.3963991855
25	0.3941284325
35	0.3947738694
50	0.4077619971
100	0.4011556675
200	0.3937565718

C) Refer Notebook

D) RMSE Errors on test set
BGD- 0.9631131
With Reg-0.93846214
With bias+reg-0.93794073

Q2)SVD:-0.89248 (RMSE error)

Q3)
NMF-0.94680129
The advantage of NMF can be obtained when we are able to see aspects, users affinity towards them and analyse their composition. As the dataset is large and the chosen k value is 346, it becomes hard to interpret any such semantic sense out of the data.

Part 2

Q 1

Categories: 28

Subcategories: 53

Q 2) 8888*68953 , Approx 5 GB (4,902,834,240 bytes)

Q3

LSI , k = 70 (Near to number of subcategories, so that we can see relations if they exist)

```
Topic 1 :
```
알래스카의, 그다음, 발생하는, 데에도, 걸리지, 겨울은, 기록이, 시작된, 따뜻하고, 해당한다,

```
Topic 2 :
school,cancer,game,student,children,polic,drug,patient,player,citi,

Topic 3 :
redston,viacom,dauman,herzer,amus,abram,sumner,compani,trust,paramount,

Topic 4 :
game,cancer,team,player,cell,season,yanke,pitsiladi,warburg,warrior,

Topic 5 :
cancer,cell,patient,warburg,tumor,drug,afghan,taliban,mutat,kenyatta,

Topic 6 :
cancer,school,cell,warburg,tumor,trump,patient,student,mutat,drug,

Topic 7 :
compani,netflix,bank,percent,market,nationstar,investor,billion,price,inve
st,

Topic 8 :
afghan,taliban,kunduz,hospit,hutchinson,afghanistan,militari,mullah,forc,f
aizullah,

Topic 9 :
pitsiladi,bekel,athlet,olymp,marathon,wada,dope,runner,russia,russian,

Topic 10 :
wada,game,dope,russia,athlet,russian,european,clinton,olymp,britain,

Topic 11 :
```

clinton,ali,sander,obama,turkey,kurdish,rhode,european,democrat,voter,

Topic 12 :
pitsiladi,clinton,bekel,sander,voter,european,vote,yanke,kenyatta,britain,

Topic 13 :
netflix,hast,sarando,clinton,wada,mccord,network,dope,stream,subscrib,

Topic 14 :
sex,clinton,polic,purdi,sander,prostitut,women,law,offic,democrat,

Topic 15 :
ali,clinton,muhammad,sander,frazier,nationstar,lonni,taliban,foreman,democrat,

Topic 16 :
ali,netflix,polic,kurdish,turkey,islam,kurd,muhammad,turkish,pitsiladi,

Topic 17 :
willerslev,ali,european,dna,sex,britain,aborigin,union,prostitut,genom,

Topic 18 :
willerslev,nationstar,clinton,dna,mortgag,kurdish,aborigin,equiti,kurd,sander,

Topic 19 :
sex,china,prostitut,women,appl,rhode,warburg,cancer,brothel,traffick,

Topic 20 :
china,volkswagen,purdi,appl,yanke,chines,disney,compani,googl,korea,

Topic 21 :
curri,warrior,karageorg,kerr,cavali,game,jame,coach,player,rugbi,

Topic 22 :
mossack,fonseca,ponsoldt,goltz,european,der,britain,polic,cancer,account,

Topic 23 :
karageorg,mossack,fonseca,ponsoldt,parasit,goltz,helminth,vik,der,purdi,

Topic 24 :
karageorg,nationstar,rhode,wrestl,mortgag,iran,saudi,webber,obama,nuclear,

Topic 25 :
purdi,rhode,china,iran,saudi,nuclear,korea,obama,nationstar,curri,

Topic 26 :
karageorg,purdi,clinton,china,disney,warburg,sex,willerslev,cancer,prostitut,

Topic 27 :
china,parasit,disney,helminth,vik,chines,hookworm,netflix,park,shanghai,

Topic 28 :
volkswagen,saudi,emiss,rhode,car,diesel,iran,casino,oil,vehicl,

Topic 29 :
volkswagen,cloud,segreg,nationstar,emiss,farragut,black,car,diesel,white,

Topic 30 :
warburg,cloud,polic,curri,warrior,saudi,islam,kosovo,clinton,parasit,

Topic 31 :
warburg,cloud,rugbi,purdi,student,cell,ambul,schoning,transcar,volkswagen,

Topic 32 :
rugbi,appl,warburg,app,schoning,googl,parasit,music,facebook,segreg,

Topic 33 :
warburg,student,curri,warrior,purdi,volkswagen,cell,mullah,ballet,transcar,

Topic 34 :
appl,republican,app,priebu,googl,rugbi,warburg,senat,rhode,hous,

Topic 35 :
volkswagen,student,nationstar,appl,app,islam,googl,colleg,kosovo,european,

Topic 36 :
saudi,kosovo,mullah,disney,islam,volkswagen,arabia,mansour,purdi,cloud,

Topic 37 :
mullah,mansour,taliban,pakistan,casino,clinton,purdi,pakistani,haqqani,rugbi,

Topic 38 :
yuliya,mullah,mansour,polic,vitali,taliban,ozick,fed,percent,pakistan,

Topic 39 :
rugbi,wada,korea,schoning,north,polic,bank,nuclear,missil,korean,

```
Topic 40 :
korea,kesha,north,missil,nuclear,korean,appl,kim,luke,djokov,

Topic 41 :
yuliya,rugbi,vitali,schoning,rhode,curri,warrior,disney,iran,volkswagen,

Topic 42 :
rhode,wada,iran,disney,polic,obama,djokov,presid,casino,panetta,

Topic 43 :
yuliya,ballet,korea,ozick,pult,vitali,north,missil,art,kim,

Topic 44 :
kesha,luke,rhode,purdi,ozick,parasit,iran,polic,charter,cancer,

Topic 45 :
ambul,transcar,patient,kesha,mullah,opioid,equiti,hors,drug,ozick,

Topic 46 :
hors,derbi,korea,ambul,transcar,north,nyquist,race,colt,saudi,

Topic 47 :
pult,hors,derbi,nyquist,colt,loan,sharehold,mckathan,duga,zika,

Topic 48 :
hors,derbi,casino,nyquist,colt,zika,mckathan,cell,race,trainer,

Topic 49 :
rousseff,hastert,phelp,bank,temer,gay,priebu,clinton,percoco,brazil,

Topic 50 :
ballet,isra,cohn,palestinian,israel,choreograph,militari,danc,hors,priebu,

Topic 51 :
gay,mateen,transgend,orlando,islam,kesha,gun,republican,disney,attack,

Topic 52 :
ballet,casino,rousseff,danc,choreograph,brazil,temer,kesha,cuban,appl,

Topic 53 :
ballet,cohn,hastert,percoco,clinton,cuomo,gay,danc,pult,hors,

Topic 54 :
rousseff,cohn,brazil,temer,art,brazilian,impeach,gay,museum,phelp,
```

```
Topic 55 :
priebu,casino,penguin,appl,music,goal,shark,collin,harvey,pittsburgh,

Topic 56 :
percoco,cohn,cuomo,karageorg,governor,gawker,facebook,cruz,campaign,blasio
,

Topic 57 :
gawker,facebook,dog,ballet,priebu,thiel,compani,tribun,harvey,collin,

Topic 58 :
djokov,priebu,appl,match,tenni,nadal,clinton,french,republican,tournament,

Topic 59 :
hastert,yanke,iraq,cohn,dog,falluja,percoco,islam,iraqi,hamilton,

Topic 60 :
percoco,cuomo,yanke,phelp,casino,kosovo,voter,dog,fed,rate,

Topic 61 :
cohn,percoco,cuomo,voter,harvey,collin,met,djokov,syndergaard,poll,

Topic 62 :
phelp,bowman,olymp,cruz,nabokov,swim,iraq,falluja,music,hamilton,

Topic 63 :
yanke,priebu,phelp,isra,palestinian,israel,disney,ryan,girardi,voter,

Topic 64 :
disney,driver,iger,shanghai,takata,priebu,nabokov,park,airbag,uber,

Topic 65 :
disney,islam,falluja,iraq,percoco,cruz,recip,iraqi,cook,food,

Topic 66 :
dog,copping,disney,appl,drug,rousseff,cruz,anim,volkswagen,warburg,

Topic 67 :
dog,cohn,phelp,hamilton,copping,bond,should,islam,been,broadway,

Topic 68 :
phelp,facebook,gawker,nationstar,driver,princ,casino,insur,warburg,opioid,

Topic 69 :
```

blast,brain,spear,aravena,murray,cruz,gay,ambul,rugbi,island,

Topic 70 :
gawker,kundla,china,hamilton,nabokov,priebu,should,perez,kosovo,peat,

Q 4) **PLSI**

Topic: 1
china, chines, compani, said, hong, govern, beij, kong, trade, year,
"INTERNATIONAL"
Topic: 2
sander, voter, clinton, poll, vote, state, democrat, primari, percent,
win,
"ELECTIONS"
Topic: 3
said, hospit, fire, bodi, report, die, peopl, death, one, day,
"HEALTH"
Topic: 4
countri, said, govern, refuge, european, nation, europ, franc, year,
intern,
"EUROPE"
Topic: 5
said, bank, israel, compani, isra, palestinian, custom, use, secur, offer,
"MIDDLE EAST"
Topic: 6
friend, want, just, didn, know, day, time, get, ask, never,
Topic: 7
said, match, open, play, tenni, djokov, player, french, tournament, year,
Topic: 8
said, attack, islam, state, kill, group, govern, forc, militari, war,
Topic: 9
appl, app, can, use, servic, like, compani, will, phone, new,
Topic: 10
britain, european, union, british, vote, leav, london, referendum, said,
europ,
Topic: 11
dog, like, one, peopl, can, even, cloud, world, make, way,
Topic: 12
film, show, movi, like, said, one, charact, star, watch, seri,
Topic: 13
hit, run, game, yanke, said, inning, pitch, first, season, two,
Topic: 14
book, girl, read, write, said, like, novel, stori, school, one,

```
Topic: 15
game, goal, first, said, play, score, second, two, team, seri,
Topic: 16
women, ali, said, sexual, sex, men, woman, peopl, femal, rape,
Topic: 17
test, said, health, zika, women, viru, state, abort, infect, case,
Topic: 18
music, song, like, one, said, album, princ, new, play, perform,
Topic: 19
state, said, investig, citi, new, mayor, york, blasio, governor, cuomo,
Topic: 20
club, spain, leagu, team, player, year, madrid, champion, one, win,
Topic: 21
parti, said, polit, presid, ryan, senat, brazil, leader, hous, former,
Topic: 22
gay, orlando, said, peopl, transgend, mateen, bathroom, shoot, nightclub,
lesbian,
Topic: 23
financi, bank, money, compani, fund, invest, billion, investor, firm,
loan,
Topic: 24
car, said, vehicl, recal, safeti, compani, model, drive, driver, takata,
Topic: 25
room, hous, home, said, hotel, one, live, two, park, open,
Topic: 26
art, work, artist, museum, said, exhibit, new, one, year, will,
Topic: 27
north, said, south, korea, volkswagen, unit, state, fire, korean, kim,
Topic: 28
research, studi, scientist, can, cancer, one, said, like, human, peopl,
Topic: 29
presid, obama, democrat, elect, hous, polit, vote, will, leader, senat,
Topic: 30
think, peopl, thing, don, know, can, get, want, like, work,
Topic: 31
law, court, rule, state, said, justic, right, feder, case, legal,
Topic: 32
school, student, educ, colleg, univers, said, class, year, teacher,
graduat,
Topic: 33
show, broadway, said, play, music, hamilton, theater, will, toni, year,
Topic: 34
compani, energi, said, deal, billion, year, will, climat, industri,
percent,
Topic: 35
```

said, water, citi, peopl, island, year, one, can, day, local,
Topic: 36
said, one, kenya, just, year, day, kenyatta, can, call, like,
Topic: 37
time, articl, new, report, news, publish, york, editor, public, newspap,
Topic: 38
trump, republican, campaign, said, candid, senat, parti, donald, support, democrat,
Topic: 39
drug, said, health, patient, care, medic, doctor, hospit, can, use,
Topic: 40
state, unit, said, american, militari, offici, nation, iran, saudi, presid,
Topic: 41
said, say, one, like, just, year, get, luke, shoe, want,
Topic: 42
food, restaur, cook, like, can, make, said, recip, one, eat,
Topic: 43
said, danc, ballet, wed, dancer, revolut, work, allen, one, cultur,
Topic: 44
percent, rate, year, economi, market, econom, increas, said, job, growth,
Topic: 45
polic, said, offic, gun, shoot, kill, arrest, shot, man, video,
Topic: 46
clinton, said, campaign, presid, hillari, speech, use, polit, attack, say,
Topic: 47
father, new, mother, son, univers, york, famili, graduat, work, daughter,
Topic: 48
board, redston, said, compani, viacom, director, execut, trust, control, lawyer,
Topic: 49
said, design, store, card, like, custom, new, can, use, shop,
Topic: 50
case, said, court, judg, lawyer, charg, investig, offic, prosecutor, prison,
Topic: 51
met, said, collin, wright, syndergaard, walker, matz, back, season, will,
Topic: 52
season, driver, uber, year, race, first, last, time, quarter, seri,
Topic: 53
team, island, cup, world, game, said, unit, argentina, state, copa,
Topic: 54
said, turkey, erdogan, turkish, one, name, marathon, can, run, like,
Topic: 55
citi, build, new, said, street, york, brooklyn, park, apart, neighborhood,

```
Topic: 56
million, price, market, insur, cost, ticket, percent, bond, new, year,
Topic: 57
facebook, twitter, media, post, social, peopl, news, said, europ,
european,
Topic: 58
said, travel, flight, airport, plane, secur, airlin, sea, passeng, island,
Topic: 59
art, million, said, auction, museum, sale, work, new, year, will,
Topic: 60
like, can, wine, one, bottl, use, look, make, get, say,
Topic: 61
fan, team, game, soccer, franc, stadium, will, match, england,
championship,
Topic: 62
said, compani, new, tax, busi, execut, program, million, plan, advertis,
Topic: 63
game, point, said, warrior, curri, jame, team, play, final, first,
Topic: 64
fashion, design, said, show, wear, will, dress, new, year, like,
Topic: 65
olymp, athlet, russian, said, russia, sport, dope, rio, puerto, game,
Topic: 66
compani, sale, said, year, busi, percent, new, execut, chief, retail,
Topic: 67
race, hors, said, one, derbi, like, track, time, year, last,
Topic: 68
black, said, church, white, state, children, peopl, counti, racial, mani,
Topic: 69
googl, said, golf, cuban, day, hole, round, cuba, one, johnson,
Topic: 70
player, team, said, coach, leagu, year, play, sport, draft, season,

Q5)
Topic 1 :
investig, compani, report, file, case, lawyer, lawsuit, email, depart,

Topic 2 :
art, work, museum, artist, exhibit, galleri, show, new, paint,

Topic 3 :
percent, price, rose, oil, stock, fell, market, energi, cent,

Topic 4 :
travel, hotel, trip, airport, guest, includ, visit, park, book,
```

Topic 5 :
film, music, movi, song, album, perform, play, festiv, new,

Topic 6 :
game, point, said, play, warrior, curri, final, player, jame,

Topic 7 :
plane, fire, flight, air, egyptian, drone, aircraft, canada, fort,

Topic 8 :
anderson, cox, nativ, medit, ami, pugh, farc, duncan, wallac,

Topic 9 :
gay, transgend, right, peopl, bathroom, bar, lesbian, gender, north,

Topic 10 :
sherman, toast, ark, duckpin, ticketmast, norwegian, bowl, oslo, twain,

Topic 11 :
food, restaur, cook, recip, chef, can, make, chicken, eat,

Topic 12 :
music, opera, orchestra, compos, concert, hall, bitcoin, met, perform,

Topic 13 :
dog, puerto, rico, cat, rican, pet, anim, island, tillman,

Topic 14 :
tribun, publish, gannett, earthquak, ecuador, newspap, mcdonald, ferro,
mitchel,

Topic 15 :
bank, financi, rate, tax, money, million, economi, econom, market,

Topic 16 :
million, auction, sale, sold, market, price, christi, dealer, buyer,

Topic 17 :
china, chines, beij, hong, kong, parti, taiwan, govern, foreign,

Topic 18 :
state, said, govern, unit, attack, militari, islam, american, offici,

Topic 19 :

skin, hair, bodi, product, beauti, makeup, pierc, vitamin, perfum,

Topic 20 :
said, year, will, new, one, also, can, state, peopl,

Topic 21 :
facebook, compani, appl, onlin, googl, technolog, app, media, use,

Topic 22 :
drug, health, medic, patient, hospit, care, doctor, insur, treatment,

Topic 23 :
said, offic, case, polic, court, charg, judg, lawyer, prosecutor,

Topic 24 :
trump, clinton, republican, campaign, democrat, parti, said, sander,
state,

Topic 25 :
princ, baker, minneapoli, willerslev, estat, musician, paisley, sri,
purpl,

Topic 26 :
game, run, said, hit, yanke, met, inning, pitch, season,

Topic 27 :
citi, new, said, york, mayor, offici, blasio, depart, peopl,

Topic 28 :
season, throne, jon, snow, game, ramsay, cersei, dragon, aravena,

Topic 29 :
new, york, graduat, univers, father, coupl, marri, receiv, mother,

Topic 30 :
design, fashion, wear, brand, dress, collect, cloth, show, men,

Topic 31 :
hastert, bedroom, room, pool, dutert, properti, philippin, bathroom, hous,

Topic 32 :
player, team, play, coach, leagu, season, game, year, said,

Topic 33 :
sex, panama, prostitut, der, buffett, worker, mossack, fonseca, offshor,

Topic 34 :
space, nasa, light, planet, earth, hole, moon, black, sun,

Topic 35 :
brazil, rousseff, presid, brazilian, parti, countri, polit, impeach, temer,

Topic 36 :
car, driver, vehicl, volkswagen, recal, test, drive, safeti, compani,

Topic 37 :
iran, iranian, rhode, nuclear, saudi, sanction, agreement, deal, tehran,

Topic 38 :
build, citi, hous, street, new, home, said, park, space,

Topic 39 :
theater, show, play, broadway, will, music, perform, award, new,

Topic 40 :
like, one, can, just, time, said, make, get, way,

Topic 41 :
said, one, famili, peopl, day, year, live, friend, time,

Topic 42 :
danc, ballet, wine, dancer, choreograph, compani, grape, vineyard, jone,

Topic 43 :
north, korea, japan, south, nuclear, obama, korean, kim, japanes,

Topic 44 :
하지만, 말했다, 것이다, falun, gong, 레이저, kasher, 때문에, leggero,

Topic 45 :
jackson, beer, hornacek, spear, drink, knick, anthoni, softe, brew,

Topic 46 :
darti, fnac, circu, conforama, peterson, wiltshir, steinhoff, hobb, nash,

Topic 47 :
compani, busi, billion, million, execut, year, invest, deal, market,

Topic 48 :

water, plant, climat, energi, environment, power, electr, product, industri,

Topic 49 :
robot, hadid, ozick, böhmermann, seneg, polanski, cashin, dakar, zaha,

Topic 50 :
team, game, player, soccer, play, goal, leagu, fan, club,

Topic 51 :
state, law, court, senat, democrat, republican, hous, bill, justic,

Topic 52 :
redston, viacom, hors, race, derbi, dauman, nyquist, amus, director,

Topic 53 :
book, women, time, articl, new, stori, write, read, wrote,

Topic 54 :
luke, kesha, opera, graham, mose, soni, haywood, abballa, widad,

Topic 55 :
rugbi, zappa, zealand, rudolph, schoning, australia, africa, dweezil, abdeslam,

Topic 56 :
show, season, televis, seri, watch, network, episod, netflix, comedi,

Topic 57 :
jewish, tiger, jew, templ, rabbi, mosqu, muslim, religi, kosher,

Topic 58 :
ali, muhammad, fight, box, ring, clay, frazier, champion, sport,

Topic 59 :
research, studi, zika, diseas, scientist, health, viru, found, brain,

Topic 60 :
india, cuban, indian, cuba, pitsiladi, kenya, german, berlin, modi,

Topic 61 :
anim, bird, speci, zoo, wildlif, gorilla, park, eleph, conserv,

Topic 62 :
school, student, univers, colleg, educ, children, class, graduat, black,

```
Topic 63 :
european, union, britain, europ, vote, british, parti, countri, leav,

Topic 64 :
olymp, russia, russian, athlet, sport, dope, rio, game, world,

Topic 65 :
allen, nepal, climb, politico, romeo, nepales, panton, juliet, cervant,

Topic 66 :
israel, isra, palestinian, gaza, tunnel, netanyahu, jerusalem, hama, bank,

Topic 67 :
church, cancer, christian, cloud, cathol, tumor, cell, andrew, pope,

Topic 68 :
game, goal, team, score, first, seri, second, play, said,

Topic 69 :
polic, shoot, gun, kill, orlando, shot, attack, peopl, victim,

Topic 70 :
island, sea, boat, water, beach, ship, ocean, fish, coast,
```

LDA has almost zero number of irrelevant words in the topics, and the topics are well formed.
As we are trying to compare the latent topics with categories/ sub categories, and every document has only one subcategory, we should keep alpha low. Similarly, we want only 10 dominant words from a topic, hence we can take beta to be low as well. Results can be manually interpreted from "LDA ALL results.pdf", where the algorithm is ran for each combination of alpha and eta.