In [0]:

```
## Reading Finance Based Papers
```

In [2]:

```
## Read the finance papers from 2016 to 2018 and store it in a file

import urllib
url = 'http://export.arxiv.org/oai2?verb=ListRecords&set=q-fin&from=2016-01-01&until=2018-11-31&metadataPref
ix=arXiv'
data = urllib.request.urlopen(url).read()

fin = open('finance1', 'wb')
fin.write(data)
```

Out[2]:

```
1853006
```

In [0]:

```
## Extract the title and abstract from papers - Read from finance1 to finance2
!xml_grep 'title|abstract' finance1  > finance2.txt
```

In [0]:

```
## Remove Junk lines , here we remove first 3 lines and last 3 lines which are not necessary
!cat finance2.txt | tail -n +4 | head -n -3 > finance3.txt
```

In [6]:

```
## Reading packages for Text classification
from sklearn import model_selection, preprocessing, linear_model, naive_bayes, metrics, svm
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn import decomposition, ensemble

import pandas, numpy, string
from keras.preprocessing import text, sequence
from keras import layers, models, optimizers
from nltk import word_tokenize
from nltk.corpus import stopwords
import sklearn
#import sklearn_crfsuite
#from sklearn_crfsuite import scorers
#from sklearn_crfsuite import metrics
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.naive_bayes import GaussianNB
from sklearn.naive_bayes import MultinomialNB
from sklearn.decomposition import TruncatedSVD
from sklearn.metrics import accuracy_score
from sklearn import metrics
```

```
Using TensorFlow backend.
```

In [7]:

```
## Stopwords import and removal
import nltk
from nltk.corpus import stopwords

nltk.download('stopwords')
stopwords = set(stopwords.words('english'))
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

In [8]:

```
# load the dataset # dataset contains combined labels and text from all training papers
data = open('labeled_sentences (1).txt').read()[:-2]
labels, texts = [], []
for i, line in enumerate(data.split("\n")):
    content = line.split()
    #print(content)
    labels.append(content[0])
    filtered_sentence = [w.lower() for w in content[1:] if not w in stopwords]
    texts.append(filtered_sentence)

# create a dataframe using texts and lables
trainDF = pandas.DataFrame()
trainDF['text'] = texts
trainDF['label'] = labels
print(trainDF['label'].unique())
trainDF.head(2)
```

```
['MISC' 'AIMX' 'OWNX' 'CONT' 'BASE']
```

Out[8]:

	text	label
0	[minimum, description, length, principle, onli...	MISC
1	[underlying, model, class, discrete,, total, e...	MISC

In [0]:

```
## Used the obtained dataset for training
train_x, valid1_x, train_y, valid1_y = model_selection.train_test_split(trainDF['text'], trainDF['label'],te
st_size=0)
```

In [10]:

```
## Convert from list to string
tempp = []

for item in train_x:
    tempp.append(" ".join(item))
#print(len(train_x))

#tempp1 =[]
#for item1 in valid_x:
    #tempp1.append(" ".join(item1))

#print(len(tempp1))

temp =[]
temp_len=0
for item2 in texts:
    temp.append(" ".join(item2))
    temp_len = temp_len+len(texts)
print(len(temp))
print(temp_len)
print(type(temp))
```

```
19162
367182244
<class 'list'>
```

In [0]:

```
# create a count vectorizer object
count_vect = CountVectorizer(analyzer='word', token_pattern=r'\w{1,}')
count_vect.fit(temp)

# transform the training and validation data using count vectorizer object
xtrain_count =  count_vect.transform(tempp)
```

In [0]:

```
## Create a classifier
import csv
trainDF2 = pandas.DataFrame()

def train_model(classifier, feature_vector_train, label, feature_vector_valid, is_neural_net=False):
    # fit the training dataset on the classifier
    #std_clf = make_pipeline(StandardScaler(with_mean=False), TruncatedSVD(100), MultinominalNB())
    #std_clf.fit(feature_vector_train, label)
    classifier.fit(feature_vector_train, label)

    # predict the labels on validation dataset
    #predictions = classifier.predict(feature_vector_valid)
    predictions = classifier.predict(feature_vector_valid)
    return predictions
    #tt = classifier.predict(feature_vector_valid)
    #labels3 = classifier.predict(feature_vector_valid)

    #trainDF2['labels'] = labels3
    #trainDF2['text']= valid_x
    #print(trainDF2)
```

In [13]:

```
## Read title and abstracts and loop through them
import re
global_list = []
title_list =[]

test = open("finance3.txt",'r').read().split("</abstract>")
#print(test[1])
for idx,i in enumerate(test):
  title = re.findall(r"(?<=<title>).*(?=</title>)",i.replace("\n",""))
  #print(title)
  abstract = re.findall(r"(?<=<abstract>).*",i.replace("\n",""))
  #print(abstract[0].replace("\n",""))
  nlist = re.split(r"(?:(?<=[^i]\.)|\.(?=[^e]))",abstract[0].replace('"',"").replace('\n',''))
  #temp_abs = re.sub(r"((?<=[^i]\.)|\.(?=[^e]))","\n",abstract[0])
  #print(abstract)
  #temp_str = temp_abs.split("\n")
  #print(temp_str[0])
  #print(nlist[1])
  global_list.append(nlist)
  title_list.append(title)
  #print(global_list)

  if idx >50:
    #print(global_list)
    break
  #print(abstract[0])
  #nlist = re.split(r"(?:(?<=[^i]\.)|\.(?=[^e]))",str(abstract))

  #print(nlist[1])

  #tempp1 =[]
  '''
  for idx, item1 in enumerate(nlist):

    if idx > 1 :
      break;
      print(item1)
      tempp1.append(" ".join(item1))
    #print(tempp1)

    xvalid_count =  count_vect.transform(tempp1)
    for item in nlist:
      print(item)
      valid_x = item
      #accuracy = train_model(naive_bayes.MultinomialNB(), xtrain_count, train_y, xvalid_count)

  '''
  #print(global_list[0])
  #print(global_list[1])
  #print(global_list[2])
  #for idx, item1 in enumerate(global_list) :
  #  if idx > 1:
  #    break
  #  print(item1)
    #tempp1.append(" ".join(item1))
    #xvalid_count =  count_vect.transform(tempp1)
    #accuracy = train_model(naive_bayes.MultinomialNB(), xtrain_count, train_y, xvalid_count)
```

```
/usr/lib/python3.6/re.py:212: FutureWarning: split() requires a non-empty pattern match.
  return _compile(pattern, flags).split(string, maxsplit)
```

In [14]:

```
## Print triples from data

#print(global_list[1])
for idx, (item, title) in enumerate(zip(global_list, title_list)):



  #print(item)
  valid_x = item
  xvalid_count =  count_vect.transform(valid_x)
  accuracy = train_model(linear_model.LogisticRegression(), xtrain_count, train_y, xvalid_count)
  #print("\n\n")
  if idx>2:
    break

  title_id = hash(str(title))
  abstract_id = hash(str(item))
  line1 = "<https://w3id.org/skg/articles/" + str(title_id) + "> <http://xmlns.com/foaf/0.1/name>" + '"' + "
 ".join(title) + '"' +"."
  line2 = "<https://w3id.org/skg/articles/" + str(title_id) + "> <http://purl.org/dc/terms/abstract> <http:/
/purl.org/dc/terms/abstract/" + str(abstract_id)+ ">"
  line3 = "<https://w3id.org/skg/articles/" + str(abstract_id) +"><http://purl.org/dc/terms/abstract/text>"
+ '"' + " ".join(item) + '"'
  print(line1,line2,line3,sep ="\n")
  for acc,element in zip(accuracy,item):
    print('<http://purl.org/dc/terms/abstract/{} > "{}"'.format(acc, element))
    #line4 = ("<http://purl.org/dc/terms/abstract/" + str(acc) + ">" + '"' + str(element) + '"' )
```

```
<https://w3id.org/skg/articles/-5159925132796795899> <http://xmlns.com/foaf/0.1/name>"Average o
ptimality for risk-sensitive control with general state space".
<https://w3id.org/skg/articles/-5159925132796795899> <http://purl.org/dc/terms/abstract> <http:
//purl.org/dc/terms/abstract/8218326234501218954>
<https://w3id.org/skg/articles/8218326234501218954><http://purl.org/dc/terms/abstract/text>"  T
his paper deals with discrete-time Markov control processes on a generalstate space  A long-run
 risk-sensitive average cost criterion is used as aperformance measure  The one-step cost funct
ion is nonnegative and possiblyunbounded  Using the vanishing discount factor approach, the opt
imalityinequality and an optimal stationary strategy for the decision maker areestablished."
<http://purl.org/dc/terms/abstract/AIMX > "  This paper deals with discrete-time Markov control
 processes on a generalstate space"
<http://purl.org/dc/terms/abstract/MISC > " A long-run risk-sensitive average cost criterion is
 used as aperformance measure"
<http://purl.org/dc/terms/abstract/MISC > " The one-step cost function is nonnegative and possi
blyunbounded"
<http://purl.org/dc/terms/abstract/OWNX > " Using the vanishing discount factor approach, the o
ptimalityinequality and an optimal stationary strategy for the decision maker areestablished."
<https://w3id.org/skg/articles/3516941775658297347> <http://xmlns.com/foaf/0.1/name>"Maximum En
tropy, the Collective Welfare Principle and the Globalization  Process".
<https://w3id.org/skg/articles/3516941775658297347> <http://purl.org/dc/terms/abstract> <http:/
/purl.org/dc/terms/abstract/6288511295907271878>
<https://w3id.org/skg/articles/6288511295907271878><http://purl.org/dc/terms/abstract/text>"  A
lthough both systems analyzed are described through two theories apparentlydifferent (quantum m
echanics and game theory) it is shown that both areanalogous and thus exactly equivalents  The 
quantum analogue of the replicatordynamics is the von Neumann equation  Quantum mechanics could
 be used toexplain more correctly biological and economical processes  It could evenencloses th
eories like games and evolutionary dynamics  We can take someconcepts and definitions from quan
tum mechanics and physics for the bestunderstanding of the behavior of economics and biology  A
lso, we could maybeunderstand nature like a game in where its players compete for a common welf
areand the equilibrium of the system that they are members  All the members of oursystem will p
lay a game in which its maximum payoff is the equilibrium of thesystem  They act as a whole bes
ides individuals like they obey a rule in wherethey prefer to work for the welfare of the colle
ctive besides the individualwelfare  A system where its members are in Nash Equilibrium (or ESS
) is exactlyequivalent to a system in a maximum entropy state  A system is stable only ifit max
imizes the welfare of the collective above the welfare of the individual If it is maximized the
 welfare of the individual above the welfare of thecollective the system gets unstable an event
ually collapses  The results ofthis work shows that the globalization process has a behavior ex
actlyequivalent to a system that is tending to a maximum entropy state and predictsthe appariti
on of big common markets and strong common currencies that willfind its equilibrium by decreasi
ng its number until they get a statecharacterized by only one common currency and only one comm
on market around theworld."
<http://purl.org/dc/terms/abstract/MISC > "  Although both systems analyzed are described throu
gh two theories apparentlydifferent (quantum mechanics and game theory) it is shown that both a
reanalogous and thus exactly equivalents"
<http://purl.org/dc/terms/abstract/OWNX > " The quantum analogue of the replicatordynamics is t
he von Neumann equation"
<http://purl.org/dc/terms/abstract/MISC > " Quantum mechanics could be used toexplain more corr
ectly biological and economical processes"
<http://purl.org/dc/terms/abstract/MISC > " It could evenencloses theories like games and evolu
```

```
tionary dynamics"
<http://purl.org/dc/terms/abstract/OWNX > " We can take someconcepts and definitions from quant
um mechanics and physics for the bestunderstanding of the behavior of economics and biology"
<http://purl.org/dc/terms/abstract/OWNX > " Also, we could maybeunderstand nature like a game i
n where its players compete for a common welfareand the equilibrium of the system that they are
 members"
<http://purl.org/dc/terms/abstract/OWNX > " All the members of oursystem will play a game in wh
ich its maximum payoff is the equilibrium of thesystem"
<http://purl.org/dc/terms/abstract/BASE > " They act as a whole besides individuals like they o
bey a rule in wherethey prefer to work for the welfare of the collective besides the individual
welfare"
<http://purl.org/dc/terms/abstract/MISC > " A system where its members are in Nash Equilibrium
(or ESS) is exactlyequivalent to a system in a maximum entropy state"
<http://purl.org/dc/terms/abstract/CONT > " A system is stable only ifit maximizes the welfare
of the collective above the welfare of the individual"
<http://purl.org/dc/terms/abstract/OWNX > "If it is maximized the welfare of the individual abo
ve the welfare of thecollective the system gets unstable an eventually collapses"
<http://purl.org/dc/terms/abstract/MISC > " The results ofthis work shows that the globalizatio
n process has a behavior exactlyequivalent to a system that is tending to a maximum entropy sta
te and predictsthe apparition of big common markets and strong common currencies that willfind
its equilibrium by decreasing its number until they get a statecharacterized by only one common
 currency and only one common market around theworld."
<https://w3id.org/skg/articles/2238203701590207870> <http://xmlns.com/foaf/0.1/name>"On the Str
ucture of General Mean-Variance Hedging Strategies".
<https://w3id.org/skg/articles/2238203701590207870> <http://purl.org/dc/terms/abstract> <http:/
/purl.org/dc/terms/abstract/-6562757750981027048>
<https://w3id.org/skg/articles/-6562757750981027048><http://purl.org/dc/terms/abstract/text>"
We provide a new characterization of mean-variance hedging strategies in ageneral semimartingal
e market  The key point is the introduction of a newprobability measure $P^{\star}$ which turns
 the dynamic asset allocationproblem into a myopic one  The minimal martingale measure relative
 to$P^{\star}$ coincides with the variance-optimal martingale measure relative tothe original p
robability measure $P$."
<http://purl.org/dc/terms/abstract/OWNX > "  We provide a new characterization of mean-variance
 hedging strategies in ageneral semimartingale market"
<http://purl.org/dc/terms/abstract/MISC > " The key point is the introduction of a newprobabili
ty measure $P^{\star}$ which turns the dynamic asset allocationproblem into a myopic one"
<http://purl.org/dc/terms/abstract/OWNX > " The minimal martingale measure relative to$P^{\star
}$ coincides with the variance-optimal martingale measure relative tothe original probability m
easure $P$."
```