Stock Sentiment Analysis using News Headlines import pandas as pd df=pd.read csv('Data.csv', encoding = "ISO-8859-1") df.head() Top1 Top2 Top3 Top4 Top5 Top6 Top7 Top8 Top16 Top17 Top18 Top19 Top20 Top21 Top22 Top23 Top24 Top25 Yeltsin's A 'hindrance to Hunters Kohl's Diana's Depleted Jack gets his **Hungry Spurs** Hughes' Chaos as Gunners so Flintoff injury The difference Sara Denver, resignation 2000-Leicester threaten Jospin landmine Russian Recovering operations': successor Scorecard instant hit skates on at Maracana builds sense rich piles on woe nurse turned caught Sold out wide of an easy between men 01-03 crusade put extracts from prevail as with new battle drawn into roulette buoys Blues pickings ice-cold Alex up for United for England solicitor opposition flattarget and women Elliott spoils E... the.. of the... scandal Tories in a panic Yet more things Irish court Campaigners Burundi peace Millennium Leader: PE points the The timing of 2000-The best Has Cubie Has Cubie Has Cubie killed On the critical halts IRA man's keep up pressure you wouldn't The main Jane bug fails to Cheerio, boyo initiative fades Scorecard German sleaze Dear doctor way forward to 01-04 lake scene recommendations killed fees? killed fees? fees? list their lives extradition to Ratcliffe know without on Nazi war after rebels re... the ECB inquiry Nor.. the ... crime... Police help United's Thatcher Pakistan Coventry Tale of Trautmann Cullinan South Raja Ideal Pinochet leaves 2000rivals on issues defence Smith lay England on the retaliate with Necaxa Real Madrid Corinthians Tony's pet Al Nassr (Saudi caught on continues his Melbourne Casablanca Holmes hospital after Useful links bears two more 01-05 before trial by down the law rack call for video (Mexico) (Spain) (Brazil) Arabia) the road project counter by Flo retellings Cape monopoly (Australia) (Morocco) show tests video at Everton of Walsh Hussain hurt by Auntie back in Putin admits Upending a Scientific Rebels Pilgrim knows McIlroy calls Leicester bin Shoaib appeal BBC worst hit Fusco wins Blair orders Lessons of 2000-United braced for How much can Thatcher fashion, even if 'shambles' but Yeltsin quit to Christmas table, Chopping evidence thwart goes to the shake-up of 3 for Irish stadium as digital TV law's hard how to iudicial review in 01-06 lays blame on facing ban Mexican wave the dress give him a you pay for... glitches a line and 'unreliable', Russian fighting spirit blueprint begins to bite extradition case failing NHS progress heart top look.. head s... Scoring ... defence claims advance e... What is Beckham Return of the **Embattled** Whingers - a From Sleepy Guardian readers: Most Blunkett the 2000-Hitches and off but Hollywood Ashes and Most wanted: Dobson attacks Doom and the Aitken released Breast cancer cane Alan Parker formidable everywhere: Hollow to outlines vision north-Gone aloft are vou all 01-07 'completely off Horlocks United screening Beyond diamonds Chloe lunettes Dome from jail 'play now, pay UDIs whingers? minority for over 11s south Greeneland survive the agenda' later'.. divide? 5 rows × 27 columns splitting dataset into training and test data train = df[df['Date'] < '20150101']</pre> test = df[df['Date'] > '20141231'] # Removing punctuations data=train.iloc[:,2:27] data.replace("[^a-zA-Z]"," ",regex=True, inplace=True) # Renaming column names for ease of access list1= [i for i in range(25)] new Index=[str(i) for i in list1] data.columns= new Index data.head(5) Out[5]: 2 9 ... 15 19 20 22 23 3 4 5 6 7 8 16 17 18 21 24 Depleted Diana s Hunters Yeltsin s A hindrance Derby raise a Southgate Kohl s Chaos as Hughes Jack gets his Leicester Hungry Spurs Gunners so Flintoff injury threaten The difference Sara Denver landmine resignation to operations Recovering glass to strikes successor Russian prevail as Scorecard instant hit Maracana builds sense rich wide of an Jospin with crusade put caught Sold out extracts from Strupar s Leeds pay drawn into roulette buoys Blues ice cold Alex up for United Elliott spoils pickings easy target for England new battle of solicitor Tories in a opposition flat the.. debut double scandal E... Burundi peace Yet more things Hopkins Irish court Campaigners The best Leader PE points the The main Has Cubie Has Cubie Has Cubie furious at Has Cubie On the critical The timing of halts IRA man initiative fades you wouldn t Scorecard German Cheerio boyo Dear doctor killed fees Nazi war crime... of Hannibal.. to Nor.. Thatcher United s Police help Pakistan Cullinan Raja Ideal Coventry issues Tale of Trautmann McGrath puts Blair Witch South Pinochet leaves Real Madrid Smith lay England on retaliate with continues his Necaxa Corinthians Tony s pet Al Nassr Saudi hospital after Useful links caught on defence bears two more India out of bandwagon Melbourne Casablanca Holmes the road down the law call for video Cape Mexico Spain Brazil project Arabia show counter by Flo before trial by retellings their misery rolls on Australia Morocco tests to Rio at Everton of Walsh monopoly video Upending a Scientific Auntie back in Hussain hurt Revenge is Putin admits Fusco wins Pilgrim knows Shoaib Thatcher McIlroy calls Leicester bin England s BBC worst hit How much table evidence Blair orders Lessons of United braced for fashion even if judicial review by shambles sweet for Christmas how to for Irish stadium appeal goes decade of as digital TV can you pay Chopping a unreliable shake up of law s hard Mexican wave the dress but lays blame jubilant give him a in extradition Russian ban fighting spirit blueprint to the top disasters for line and defence failing NHS heart progress look... Cronje head s... advance Scoring ... claims Return of the Embattled What is Beckham Guardian readers Whingers a Thuggery Most cane From Sleepy Blunkett Dobson the Breast cancer Alan Parker Hitches and off but Hollywood Ashes and Most wanted Doom and the Aitken released Gone aloft Alan Parker are you all formidable Toxins and everywhere completely Hollow to attacks play north Horlocks United screening Beyond diamonds part two Chloe lunettes Dome from jail UDIs whingers minority Ties off the Greeneland now pay later south survive agenda divide 5 rows × 25 columns # Convertng headlines to lower case for index in new Index: data[index]=data[index].str.lower() data.head() Out[6]: 0 2 3 4 5 9 ... 15 17 18 19 20 21 22 23 24 8 16 depleted hunters diana s yeltsin s a hindrance kohl s derby raise a southgate flintoff injury the difference hughes jack gets his chaos as leicester hungry spurs gunners so threaten sara denver landmine resignation to operations glass to strikes leeds successor russian recovering maracana builds sense rich caught instant hit skates on at prevail as wide of an piles on woe jospin with between men nurse turned crusade put sold out extracts from strupar s pay the drawn into roulette new battle of opposition flat buoys blues ice cold alex up for united elliott spoils pickings easy target for england and women solicitor tories in a debut double penalty scandal the... panic e... hopkins yet more things irish court campaigners the best leader burundi peace pe points the has cubie has cubie furious at the main has cubie has cubie on the critical the timing of halts ira man keep up jane you wouldn t initiative fades lake cheerio boyo dear doctor bug fails scorecard german way forward killed fees recommendations killed fees killed fees killed fees foster s lack their lives s extradition pressure on ratcliffe know without scene sleaze inquiry after rebels re... to the ecb to bite of hannibal... to nor... nazi war crime.. thatcher pakistan united s police help cullinan coventry issues tale of trautmann mcgrath puts blair witch south raja ideal pinochet leaves smith lay continues his real madrid corinthians rivals on england on retaliate with tony s pet al nassr saudi defence casablanca hospital after useful links caught on bears two more india out of bandwagon melbourne necaxa mexico holmes arabia the road down the law the rack call for video cape spain brazil project before trial by retellings counter by flo their misery rolls on australia morocco show tests to rio at everton of walsh monopoly video auntie back in hussain hurt revenge is putin admits upending a scientific fusco wins rebels england s blair orders pilgrim knows mcilroy calls leicester bin shoaib appeal bbc worst hit how much lessons of united braced for fashion even if sweet for christmas table chopping judicial review by shambles yeltsin quit to evidence how to for irish stadium goes to the decade of as digital tv can you pay shake up of but lays blame jubilant unreliable mexican wave the dress give him a glitches a line and in extradition russian fighting spirit ban blueprint top disasters begins to bite failing nhs heart progress look.. on e... cronje head s... scoring ... defence claims case advance return of the embattled what is beckham dobson guardian readers whingers a thuggery from sleepy blunkett the most cane off but alan parker hitches and breast cancer hollywood ashes and most wanted aitken released gone aloft formidable north alan parker are you all toxins and everywhere completely hollow to outlines vision attacks play chloe lunettes from jail horlocks united screening beyond diamonds part two whingers minority udis off the greeneland for over s now pay later south survive divide agenda 5 rows × 25 columns #Merging all the columns headlines = [] for row in range(0,len(data.index)): headlines.append(' '.join(str(x) for x in data.iloc[row, 0:25])) headlines[0] 'a hindrance to operations extracts from the leaked reports scorecard hughes instant hit buoys blues jack gets his skates on at ice cold alex chaos as maracana builds up for united depleted leicester prevail as elliott spoils everton s party hungry spurs sense rich pickings gunners so wide of an easy target derby raise a glass to strupar s debut double southgate strikes leeds pay the penalty hammers hand robson a youthful lesson saints party like it wear wolves have turned into lambs stump mike catches testy gough s taunt langer escapes to hit flintoff injury piles on woe for england hunters threaten jospin with new battle of the somme kohl s successor drawn into scandal the difference between men and women sara denver nurse turned solicitor diana s landmine crusade put tories in a panic yeltsin s resignation caught opposition flat footed russian roulette sold out recovering a title' from sklearn.feature extraction.text import CountVectorizer from sklearn.ensemble import RandomForestClassifier ## implement BAG OF WORDS countvector=CountVectorizer(ngram range=(2,2)) traindataset=countvector.fit transform(headlines) In [11]: # implement RandomForest Classifier randomclassifier=RandomForestClassifier(n estimators=200,criterion='entropy') randomclassifier.fit(traindataset,train['Label']) RandomForestClassifier(criterion='entropy', n estimators=200) In [12]: ## Predict for the Test Dataset test transform= [] for row in range(0,len(test.index)): test transform.append(' '.join(str(x) for x in test.iloc[row,2:27])) test dataset = countvector.transform(test transform) predictions = randomclassifier.predict(test dataset) In [13]: ## Import library to check accuracy from sklearn.metrics import classification\_report,confusion\_matrix,accuracy\_score matrix=confusion matrix(test['Label'], predictions) print (matrix) score=accuracy score(test['Label'], predictions) report=classification report(test['Label'], predictions) print(report) [[143 43] [ 8 184]] 0.8650793650793651 precision recall f1-score support 0.95 0.77 0.85 186 0.81 0.88 192 0.96 0.87 378 accuracy 0.88 0.86 0.86 378 macro avg 0.88 0.87 0.86 weighted avg In [15]: import seaborn as sns import matplotlib.pyplot as plt ax = sns.heatmap(matrix, annot=True, cmap='Blues') ax.set title('Seaborn Confusion Matrix with labels\n\n'); ax.set xlabel('\nPredicted Values') ax.set ylabel('Actual Values '); ## Ticket labels - List must be in alphabetical order ax.xaxis.set ticklabels(['False','True']) ax.yaxis.set ticklabels(['False','True']) ## Display the visuablization of the Confusion Matrix. plt.show()

False True
Predicted Values

Here i have used bag of word word embedding technique and random forest aglortihm for more good accuracy we can use pre trained word embedding model like word2vec and nueral network

Seaborn Confusion Matrix with labels

43

1.8e+02

1.4e+02

160

- 140 - 120 - 100