



News Articles Sorting





OBJECTIVE

A Machine Learning NLP Model for predicting the category of news with the help of logistic regression classification method.

Dataset..

Dataset contains the training and testing part.
Having id , text , category column in train part
.We have to predict the category in testing
part

< BBC News Train.csv (3.2 MB)

Detail	Compact	Column
ArticleId	Text	Category
1833	worldcom ex-boss launches defence lawyers defending former worldcom chief bernie ebbers against a ba...	business
154	german business confidence slides german business confidence fell in february knocking hopes of a sp...	business
1101	bbc poll indicates	business

ArticleId	Text	Category
1833	worldcom ex-boss launches defence lawyers	business
154	german business confidence slides german	business
1101	bbc poll indicates economic gloom citizen	business
1976	lifestyle governs mobile choice faster	tech
917	enron bosses in \$168m payout eighteen	business
1582	howard truanted to play snooker conse	politics
651	wales silent on grand slam talk rhys will	sport
1797	french honour for director parker british	entertainment
2034	car giant hit by mercedes slump a slump	business
1866	rockers fuel festive film chart comedy n	entertainment
1683	blair rejects iraq advice calls tony blair h	politics
1153	housewives lift channel 4 ratings the de	entertainment
1028	uk coal plunges into deeper loss shares	business
812	bp surges ahead on high oil price oil gia	business
707	ireland 21-19 argentina an injury-time d	sport
1588	wenger signs new deal arsenal manager	sport
342	u2 s desire to be number one u2 who h	entertainment
486	hantuchova in dubai last eight daniela h	sport
1344	melzer shocks agassi in san jose second	sport
1552	moving mobile improves golf swing a m	tech
1547	hewitt overcomes wobble in sydney lle	sport
177	carry on star patsy rowlands dies actress	entertainment

Choosing Machine learning model with logistic regression (97% accuracy)

```
[19:25:39] WARNING: ..\src\learner.cc:1061: Starting in XGBoost 1.3.0, the default evaluation metric used with the objective
ulti:softprob' was changed from 'merror' to 'mlogloss'. Explicitly set eval_metric if you'd like to restore the old behavior.
[('Logistic Regression', 0.9785522788203753), ('XGboost', 0.9624664879356568), ('SVlinear', 0.9490616621983914), ('Random Fore:
t', 0.9115281501340483), ('naiveBayes', 0.900804289544236), ('Decision tree', 0.839142091152815), ('SVrbf', 0.739946380697051),
('KNN', 0.675603217158177)]
```

```
# datatest["Pr"]=a.inverse_transform(e) # prediction of test dataset goes to PredictedTest.csv file
```

```
# datatest["Pr"]=a.inverse_transform(e) # prediction of test dataset goes to PredictedTest.csv file
```

Whole code

```
In [9]: import pandas as pd      #pandas Library for data frame
        datatrain=pd.read_csv(r"C:\Users\welcome\Desktop\BBC News Train.csv")  #traindata
        datatest=pd.read_csv(r"C:\Users\welcome\Desktop\BBC News Test.csv")    #test data
        import re
        import nltk
        nltk.download('stopwords')
        from nltk.corpus import stopwords
        from nltk.stem.porter import PorterStemmer
        corpus = []
        corpust = []
        for i in range(0, 1490):
            text = re.sub('[^a-zA-Z]', ' ', datatrain['Text'][i])
            text = text.lower()
            text = text.split()
            ps = PorterStemmer()
            all_stopwords = stopwords.words('english')
            all_stopwords.remove('not')
            text = [ps.stem(word) for word in text if not word in set(all_stopwords)]
            text = ' '.join(text)
            corpus.append(text)
            corpust.append(text)
        for i in range(0, 735):
            textt = re.sub('[^a-zA-Z]', ' ', datatest['Text'][i])
            textt = textt.lower()
            textt = textt.split()
            ps = PorterStemmer()
            all_stopwords = stopwords.words('english')
            all_stopwords.remove('not')
            textt = [ps.stem(word) for word in textt if not word in set(all_stopwords)]
            textt = ' '.join(textt)
            corpust.append(textt)
        from sklearn.feature_extraction.text import CountVectorizer      #feature extraction
        cv = CountVectorizer()
        X = cv.fit_transform(corpus).toarray()      #fit and transform features in train data
        x = cv.transform(corpust).toarray()         # transform features of test data
        y = datatrain.iloc[:, -1].values
        from sklearn.preprocessing import LabelEncoder      # Label for category
        a=LabelEncoder()
        Y=a.fit_transform(y)
        from sklearn.linear_model import LogisticRegression    #Logistic regression for class
        lg=LogisticRegression(random_state=0).fit(X,Y)
        e=lg.predict(x)
        print(a.inverse_transform(e)) #predicting

        # datatest["Pr"]=a.inverse_transform(e) # prediction of test dataset goes to Finalt
        # datatest.to_csv("Finalt.csv",index=False)
```

Picking the news articles from web and predicting there types

IND vs ENG Live: Ollie Pope played a brilliant 81-run knock.

England, 4th
1: Shardul
ne Bright
other Indian
bacle,
mrah Strikes

England were dismissed for 290 as Ollie Pope and Chris Woakes scored fifties to help their team take a crucial 99-run lead on Day 2 of the 4th Test at The Oval. Pope scored 81 while Woakes played a blistering knock of 50 runs to propel England after a shaky start. Umesh Yadav took three wickets while Ravindra Jadeja claimed two to keep India in the hunt. Moeen Ali and Jonny Bairstow also added important runs along with Pope lower down the order. India will be looking for a good start from openers Rohit Sharma and KL Rahul in the second in

Copy
Copy link to highlight

```
In [13]: #for input
ap=[]
sentence = re.sub('[^a-zA-Z]', ' ', str(input()))
sentence = sentence.lower()
sentence = sentence.split()
ps = PorterStemmer()
all_stopwords = stopwords.words('english')
all_stopwords.remove('not')
sentence = [ps.stem(word) for word in sentence if not word in set(all_stopwords)]
sentence = ' '.join(sentence)
ap.append(sentence)
st=cv.transform(ap).toarray()
print(a.inverse_transform(lg.predict(st)))
```

England were dismissed for 290 as Ollie Pope and Chris Woakes scored fifties to help their team ay 2 of the 4th Test at The Oval. Pope scored 81 while Woakes played a blistering knock of 50 r
aky start. Umesh Yadav took three wickets while Ravindra Jadeja claimed two to keep India in th
stow also added important runs along with Pope lower down the order. India will be looking for
Sharma and KL Rahul in the second innings

```
['sport']
```




Mumbai Police performs Bella Ciao as Money Heist 5 releases Vol 1.

Money Heist has finally released the first volume of its fifth season on Friday. While fans have already dived into binging the episodes, Mumbai Police took the internet by surprise by releasing an instructional video showing the police performing the song Bella Ciao.

Copy

Ctrl+C

Copy link to highlight

Search Google for "Mumbai Police performs Bella Ciao as Money Heist..."

ish


```
In [15]: #for input
ap=[]
sentence = re.sub('[^a-zA-Z]', ' ', str(input()))
sentence = sentence.lower()
sentence = sentence.split()
ps = PorterStemmer()
all_stopwords = stopwords.words('english')
all_stopwords.remove('not')
sentence = [ps.stem(word) for word in sentence if not word in set(all_stopwords)]
sentence = ' '.join(sentence)
ap.append(sentence)
st=cv.transform(ap).toarray()
print(a.inverse_transform(lg.predict(st)))
```

Mumbai Police performs Bella Ciao as Money Heist 5 releases Vol 1. WEB SERIES Money Heist 5: Mumbai Police as 'I erforms Bella Ciao, watch By HT Entertainment Desk PUBLISHED ON SEP 03, 2021 08:46 PM IST On Friday, as Netflix i isodes of Money Heist, Mumbai Police dropped a video in which their band performed Bella Ciao. Watch it here. M finally released the first volume of its fifth season on Friday. While fans have already dived into binging the i i Police took the internet by surprise by releasing an instrumental version of the song Bella Ciao, featured in i w.

HT 'entertainment'



NOTE

A submission on kaggle platform score this model with 0.97551

InClass Prediction Competition

BBC News Classification

News Articles Categorization

9 teams · 2 years ago

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#)

[My Submissions](#)

[Late Submission](#)

...

Your most recent submission

Name
Finaltt.csv

Submitted
3 hours to go

Wait time
1 seconds

Execution time
0 seconds

Score
0.97551

Complete

[Jump to your position on the leaderboard](#)



Fill a direct file

```
datatest["Category"]=a.inverse_transform(e) # prediction  
datatest.to_csv("Finalt.csv",index=False)
```

A	B	C
ArticleId	Text	Category
1018	qpr keeper day heads for preston queens park rang	sport
1319	software watching while you work software that ca	tech
1138	d arcy injury adds to ireland woe gordon d arcy has	sport
459	india s reliance family feud heats up the ongoing p	business
1020	boro suffer morrison injury blow middlesbrough m	sport
51	lewsey puzzle over disallowed try england s josh le	sport
2025	blair blasts tory spending plans tony blair has launc	politics
1479	former ni minister scott dies former northern irelan	politics
27	career honour for actor dicaprio actor leonardo dic	entertainment
397	tsunami to hit sri lanka banks sri lanka s banks fac	business
1644	us economy still growing says fed most areas of the	business
263	digital uk driven by net and tv the uk s adoption of	tech
765	blunkett tells of love and pain david blunkett has s	politics
2134	ibm puts cash behind linux push ibm is spending \$1	tech
297	cage film s third week at us top nicolas cage movie	entertainment
1712	souness backs smith for scotland graeme souness k	sport
1631	ukip mep attacked german empire a uk independe	politics
942	cheaper chip for mobiles a mobile phone chip whic	tech
1549	usher leads soul train shortlist chart-topping r&b st	entertainment
516	india to deport bollywood actress india has order	entertainment
2215	bush budget seeks deep cutbacks president bush h	business
531	labour s core support takes stock tony blair has tol	politics
1541	hodges announces rugby retirement scarlets and u	sport
1340	latin america sees strong growth latin america s ec	business

For any query mail:- vaibhavchopra1408@gmail.com

Vaibhav Chopra