

HINDI SENTIMENT ANALYSIS

Submitted by

**NITIN KUMAR [RA2011047010029]
UMANG KUMAR [RA2011047010047]**

Under the guidance of
Mrs. J. JEYASUDHA

(Assistant Professor, Department of Computational Intelligence)

in partial fulfillment for the award of the degree
Of
BACHELOR OF TECHNOLOGY
in
COMPUTATIONAL INTELLIGENCE
of
FACULTY OF ENGINEERING AND TECHNOLOGY



**S.R.M. Nagar, Kattankulathur, Chengalpattu District
NOVEMBER 2022**

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this project report titled “**HINDI SENTIMENT ANALYSIS**” is the bonafide work of “**NITIN KUMAR [RA2011047010029], UMANG KUMAR [RA2011047010047]**”, who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Mrs. J. JEYASUDHA

GUIDE

Assistant Professor

Dept. of Computational Intelligence

SIGNATURE

Dr. ANNIE UTHRA

HEAD OF THE DEPARTMENT

Dept. of Computational Intelligence

Signature of the Internal Examiner

Signature of the External Exam

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my guide, **Mrs. J. JEYASUDHA**, his valuable guidance, consistent encouragement, personal caring, timely help and providing me with an excellent atmosphere for doing the project. All through the work, in spite of his busy schedule, he has extended cheerful and cordial support to me for completing this project work.

NITIN KUMAR
UMANG KUMAR


```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
```

```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
import tensorflow as tf
import random
```

```
from google.colab import files
```

```
uploaded=files.upload()
```

 No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving Sarcasm_Hindi_Tweets-SARCASTIC.csv to Sarcasm_Hindi_Tweets-SARCASTIC.csv

```
data=pd.read_csv("Sarcasm_Hindi_Tweets-SARCASTIC.csv")
```

```
data
```

	username	acctdesc	location	following	followers	totaltweets	usercreatedts	tweetcreatedts	retweetcount	text	hashtags
0	Lokeshk67004736	NaN	NaN	632	380	14186	2020-02-26 16:57:31	2020-06-23 01:50:13	82	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	[]
1	prayagraj_AU	सत्यमेव जयते\n\nभगवा ध्वज भारत का ऐतिहासिक एवं...	Uttar Pradesh, India	208	192	677	2020-06-16 10:04:04	2020-06-23 00:46:17	82	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	[]
2	brand_dbm1	संस्थापक #भगवारक्षकटीम राम राम दोस्तो 🙏 ...	New Delhi, India	1029	3027	2908	2014-11-16 17:44:40	2020-06-23 00:16:33	82	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	[]

```
data.drop(columns = data.columns[0], axis = 1, inplace= True)
```

```
data.columns = ['username', 'acctdesc', 'location', 'following', 'followers', 'totaltweets', 'usercreatedts', 'tweetcreatedts', 'retweetcount', 'text', 'hashtags']
data.head()
```

	username	acctdesc	location	following	followers	totaltweets	usercreatedts	tweetcreatedts	retweetcount	text	hashtags
0	NaN	NaN	632	380		14186	2020-02-26 16:57:31	2020-06-23 01:50:13	82	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	[]
1	सत्यमेव जयते\n\nभगवा ध्वज भारत का ऐतिहासिक एवं...	Uttar Pradesh, India	208	192		677	2020-06-16 10:04:04	2020-06-23 00:46:17	82	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	[]
2	संस्थापक #भगवारक्षकटीम राम राम दोस्तो 🙏 ...	New Delhi, India	1029	3027		2908	2014-11-16 17:44:40	2020-06-23 00:16:33	82	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	[]
-	हिन्दू हूँ तो हिंदुत्व की बात ही करूंगा।जय	2017-08-20	2020-06-22	...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	...

```
data = data.drop(['acctdesc', 'location', 'following', 'followers', 'totaltweets', 'usercreatedts', 'tweetcreatedts', 'retweetcount', 'hashtags'], axis=1)
data.head()
```

	username	text
0	NaN	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
1	सत्यमेव जयते\n\nभगवा ध्वज भारत का ऐतिहासिक एवं...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
2	संस्थापक #भगवारक्षकटीम राम राम दोस्तो 🙏 ...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
3	हिन्दू हूँ तो हिंदुत्व की बात ही करूंगा।जय	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
4	हम जिहादियों वामपंथियों कांग्रेसियों सपाइयों क...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...

```
data.username.unique()
```

```
array([nan,
       'सत्यमेव जयते\n\nभगवा ध्वज भारत का ऐतिहासिक एवं सांस्कृतिक ध्वज है । यह हिन्दुओं के महान प्रतीकों में से एक है।',
       'संस्थापक #भगवारक्षकटीम राम राम दोस्तो 🙏 🙏 🙏\n\nनोटिफिकेशन on करे',
       ...])
```

'हिन्दू हैं तो हिंदुत्व की बात ही करूंगा।जय श्री राम▶▶▶',
 'हम जिहादियों वामपंथियों कांग्रेसियों सपाइयों की कह के लेते हैं...!',
 'सर्व परवश दुःख सर्वमात्मवशं सुखम्।\nEverything that is in other's control is painful. All that is in self-control is happiness.',
 'मजबूर, मजदूर, लाचार और किसानों पर सच्चाई से लिखता हूँ। जो दिखता है:- वही आवाज उठाता हूँ। #सत्यमेव जयते\n#जय_हिन्द',
 'Alhamdulillah♥\nINDIANIN\nTaken 17-07-2018\nBy\nMdIrshad09 ♥♥♥🌺🌺',
 'हिंदू ऋषिकेश पाठक का बैकअप अकाउंट....', 'JAI SHREE RAM',
 'मण्डलअध्यक्ष चिनहट मण्डल लखनऊ ,,,मुझे भारतीय होने का गर्व है @अपने प्रधानमंत्री श्री #मोदीजी का प्रबल समर्थक हूँ',
 'पधारो म्हारे सांड. 🐘🐘\nलालूजी से चारा उधार रखें हैं! 😊\nमेरा गोबर like में मिलेगा.👍',
 'Proud Indian, Proud Hindu, Aaptards, Khangressi Doormats, Psuedosickulars, Libtards are traitors of my motherland, So they should stay away from my TL',
 'Dreamer\nLearner\nBoy with memes and dreams',
 'Founder at Been There, Done That!',
 'थोड़ी मेहनत मेरे हाथ की ! बाकी सब कृपा भोलेनाथ की। \nबम बम भोले',
 '25May1980', 'flow like a river.\n\nनिहायती बेशर्म\n\n100% satire',
 'प्रभु जी सदा ही कृपा हमपे बनाए रखना भूल कर के भी कोई भूल ना हो 🙏😊बीजेपी समर्थक देशहित सर्वोपरि 🇮🇳भारत माता की जय',
 'हम नहीं बोलते हिन्दुस्तान बोलता है,\nब्रह्मणों से ब्रह्मांड डोलता है।\n#Engineer(Ministry of Railways)#GovtOfIndia. #TelecomDepartment. #ProudBihari.',
 'तत्त्वमसि(तत् त्वम् असि) | सिफर | Nothingness | \nराजनीति में रुचि नहीं है !',
 '"Proud of Indian IN/ #Team_SCB/ #NationFirst/ My Heroes Standing on Borders/ Proud Son of Farmer/ Ardent Fan of #IndianArmy/ #TeamIndia/ RT don't endorsement",
 'video creator |\nI take jokes very seriously |\nwanna be influencer 😊\nInstagram-@Gup_Shup_queen',
 'अध्यक्ष :अखिल भारतीय पत्नी पीड़ित संघ 🇮🇳#RED #TPN #BRT',
 'Don't compare me to other girls. There's no competition. I'm one of a kind.' 😊!! Passionate for listening 🎧 music 🎵',
 'IN RT पर सहमति नहीं!',
 'हम नहीं इतिहास बोलता है ब्राह्मणों से ब्रह्माण्ड डोलता है 🙏🙏▶▶▶',
 'ना कोई मोह है ना कोई माया | बेबाक | बेफिक्र | बेपरवाह | बेसुध | बेसहारा | बेईमान | बेहतर |',
 '#mahadev bhakt #gattaparty sadasya',
 '▶ राष्ट्रहित▶ सर्वोपरि ▶ राष्ट्रवादी IN', 'PROUD TO BE INDIAN',
 'ये राजनीति का अखाड़ा है मियां यहां शब्दों से पटकनी दी जाती है (उत्तरप्रदेश)~ (Gorakhpur) ▶ देशद्रोहों दूर रहे @Narendramodi Or @Myogiadityanath Fan IN',
 'इन्हीं बिगड़े दिमागों में, अज़ाब खुशियों के लच्छे है, हमें पागल ही रहने दो, की हम पागल ही अच्छे है !!',
 'खाक भी जिस जमीं का पारस है ,शहर मशहूर वो बनारस है। यही खुदपंसदगी बनारस का यथार्थ है। जो सब कुछ अपने ठेंगे पर रखती है।',
 '#Maldives▶ My country\nis first for me 🇮🇳\n▶ #जय_श्री राम 🇮🇳 love #IndianArmy \n\n सह संचालक #BRT',
 '▶ जब तक माथे पे सजी हैं \nभस्म की ढाल\nबाल न बाका कर सके \nजन्म जन्म तक काल ▶\n 🙏 जय जय श्री महकाल 🙏',
 'देशभक्त 🇮🇳\nहिन्दुस्तान और हिन्दुत्व से बढ़कर कुछ भी नहीं♥\nINजय हिन्दINजय हिंदुत्वIN 🙏🇮🇳 जय माता दी 🇮🇳भारतीयININ 🙏 Respect Indian Forces 🇮🇳',
 'हम पंछी उन्मुक्त गगन के',
 'कुछ दलों को यह डर सताता है कि सामाजिक वर्ग जातियों का भेद भुलाकर एकजुट हो गए तो उनकी राजनीति की दुकान कैसे चलेगी?',
 'आप #मुसलमान हो इसलिए मैं #हिंदू हूँ, अन्यथा मैं तो #विश्वमानव हूँ (वीर सावरकर)\n#जय_श्री राम स्वीट दीदी @Real_jyoti',
 'बुरबादी ही है जब मुकद्दर मेरा,फर्क नहीं पड़ता कि फिर तू है या कोई और.... 🙏 सम्मान की इच्छा नहीं/अपमान का भय नहीं 🙏🙏',
 'suspend 3 twitter account, patriotic 🇮🇳\n\nOnly Nationalism 🇮🇳\n\nHindi, Hindustan ♥',
 '@akshaykumar और साउथ स्टार @urstrulyMahesh फैन.... \n\nमुस्कराता चेहरा , इतिहास गहरा , \n No DM \n..... जो जैसा बोलेंगा , वो वैसा सुनेगा 🙏 ▶▶▶',
 'Nation firstIN ▶ Nature lover 100%follow back',
 '♥🌺 शाम का आँचल ओढ़ के आयी देखो वो रात सुहानी आ लिख दें हम दोनों मिलके अपनी ये प्रेम कहानी हम रहें या ना रहे ♥🌺 my tweet in ♥\n#मितवा #साथी_संग_संगिनी #दूरिया 😊',
 'विज्ञान के गुरुजी (गणित) \n\n इस दुनिया में कुछ भी असानी से नहीं मिलता उसके लिए हमें मेहनत करनी पड़ती है',
 'त्याग दो सब ख्वाहिशें कुछ अलग करने के लिए,\n'राम' ने खोया बहुत कुछ'श्री राम' बनने के लिए।",
 'नागेन्द्रहाराय त्रिलोचनाय\nभस्माङ्गरागय महेश्वराय ।\nनित्याय शुद्धाय दिगम्बराय\nतस्मै नमः शिवाय !',
 'Nation First ▶ Nature Lover, 100% Follow back ,, Ye Backup Account hai .\n main Account 🙏 @KingAbhi144',
 'बेटी पढ़ाओ बेटी बचाओ',
 'Save Animal & Tree, Save World 🇮🇳 🇮🇳 Nationalist 🇮🇳',
 'हर हर महादेव 🙏 ▶ RT सहमति नहीं❌',
 'चम्बल मुरैना बाला : बेरोजगार : जज्बाती : ... \n\nज़रूरत है मुझे कुछ नए नफरत करने वालों की...! \nपुराने वाले तो अब चाहने लगे हैं मुझे...!!',
 'कमल भी हूँ केसरी भी, मिलना किससे है तय आप करें!\nEx.MLA, #Pilot, #Writer, #Blogger, #Traveller, #Mountanieer, #Photographer, \nNone of these !! 😊',
 'heart ♥ belongs to #बागीबलिया IN | IndiaWings',
 'पक्का भक्त डांस बार विरोधी बार बाला आलोचक',
 'नाम और पहचान भले छोटी हो मगर खुद की हो🙏\n\nधृति: क्षमा दमोस्\u200dदतयं शौचमिन्\u200dदयिनिग्रहः।\nधीर्विद्या सत्\u200dदयमक्रोधो दशकं धर्मलक्षणम्\u200dद॥',
 '#टीम_पुं_नेतन्याहु\n\n Landed on 24 December🛩️ 🙏\n\n Love photography 📷 Sports 🏆 Dream:-Reach to goal 🏆 & no.1 yari 🙏\n\n Trust to be for special one 🙏',
 '🚫\nmy personal account 🙏😊 \n\nIndian Army ♥\nजय जवान जय किसान 🙏',
 'कट्टर हिन्दू हूँ हिंदुत्व की बातें करता हूँ । हिन्दुस्तानी हूँ हिंदु राष्ट्र की बातें करता हूँ ।आतंकवादी ,चमचे ,खान , स्टार किड्स,जौहर को फॉलो करने वाले दुर रहे',

```
data.isnull().sum().sum()
```

```
700
```

```
data.fillna(value = 0,
            inplace = True)
```

```
data.head()
```

	username	text
0	0	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
1	सत्यमेव जयते\n\nभगवा ध्वज भारत का ऐतिहासिक एवं...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
2	संस्थापक #भगवारक्षकटीम राम राम दोस्तो 🙏 ...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
3	हिन्दू हूँ तो हिंदुत्व की बात ही करूँगा।जय श्र...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...
4	हम जिहादियों वामपंथियों कांग्रेसियों सपाइयों क...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...

```
def get_tweets(dataset):
    tweets = [x for x in dataset['text']]
    labels = [x for x in dataset['username']]
    return tweets, labels
```

```
tweets, labels = get_tweets(data)
```

```
tweets[0], labels[0]
```

```
('हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ले के बच्चों को अपने स्कूल आने को कहते थे। जब मैं बड़ा हुआ तो पता चला- स्कूल मास्टर के बेटे का, बस उनके भतीजे की, कैटीन उनके भांजे की और पढ़ाने वाले भी सब उनके रिश्तेदार ही थे..! \n\nआज मास्टर जी में सरकार में मंत्री हैं..! \n\n#कटाक्ष',
0)
```

```
def text_prepare(input_text, ar_text):
    out_text = delete_links(input_text)
    out_text = delete_repeated_characters(out_text)
    out_text = clean_text(out_text)
    out_text = delete_stopwords(out_text)
    out_text = out_text.lower()
    return out_text
```

```
import re
```

```
def delete_links(input_text):
    pettern = r'''(?i)\b((?:https?://|www\d{0,3}[.]|[a-z0-9.\-]+[.][a-z]{2,4}/)(?:[^\s()<>+|\\((^\\s()<>+|\\((^\\s()<>+\\))\\))+(?:\\((^\\s()<>+|\\((^\\s()<>+\\))\\))\\)|[^\s`!()\\[\\
```

```
out_text = re.sub(pettern, ' ', input_text)
return out_text
```

```
def delete_repeated_characters(input_text):
    pattern = r'(\.){1,2}'
    out_text = re.sub(pattern, r'\1', input_text)
    return out_text
```

```
def clean_text(input_text):
    replace = r'[/(){} \[\]|@â& ;\?\'\"*...:~+&!@!@-?]'
    out_text = re.sub(replace, " ", input_text)
    words = nltk.word_tokenize(out_text)
    words = [word for word in words if word.isalpha()]
    out_text = ' '.join(words)
    return out_text
```

```
def delete_stopwords(input_text):
    stop_words = set(nltk.corpus.stopwords.words("arabic") + nltk.corpus.stopwords.words("english"))
    tokenizer = nltk.tokenize.WhitespaceTokenizer()
    tokens = tokenizer.tokenize(input_text)
    wnl = nltk.WordNetLemmatizer()
    lemmatizedTokens=[wnl.lemmatize(t) for t in tokens]
    out_text = [w for w in lemmatizedTokens if not w in stop_words]
    out_text = ' '.join(out_text)
    return out_text
```

```
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

```
import nltk
nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
True
```

```
import nltk
nltk.download('omw-1.4')
```

```
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
True
```



```
import nltk
nltk.download('stopwords')

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
True
```

TOKENIZATION

```
data['Processed Text']=data['text'].apply(text_prepare, args=(False,))
data.sample(10)
```

	username	text	Processed Text
997	जय महाकाल 🙏🙏\nचिंतक । विचारक । लेखक । राजनितिक...	अगर आज नेहरू जी होते तो शी जिनपिंग की बीवी को ...	अगर आज तरह वश कर और पर कर और सब द जय आक
5159	हर उस व्यक्ति का मैं समर्थन करता हु ! जिसके दि...	अगर आज नेहरू होते तो, शी जिनपिंग की बीवी को एड...	अगर आज तरह वश कर और पर कर और सब द आक
5252	▶▶ जय श्री राम जय श्री राम ▶ हर हर महादेव 🌺 ♫ हर...	अगर आज नेहरू जी होते तो शी जिनपिंग की बीवी को ...	अगर आज तरह वश कर और पर कर और सब द जय आक
505	भारत विरोधी, हिन्दू विरोधी दूर रहें 🙏 जो पढ़ा...	अगर आज नेहरू जी होते तो शी जिनपिंग की बीवी को ...	अगर आज तरह वश कर और पर कर और सब द जय आक
1765	ना कविता लिखता हूँ ना कोई छंद लिखता हूँ अपने आ...	Wah gobhi ji wah ! #मोदी_ही_पप्पू है #FactsMat...	wah gobhi ji wah factsmatter hypocrite liarinc...
2036	We are friends if you can communicate in Morse.	'Bahar ki gandgi mat khaya kar jo kahana hai m...	bahar ki gandgi mat khaya kar jo kahana hai ma...
1858	कुछ दलों को यह डर सताता है कि सामाजिक वर्ग जात...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	एक जब बस और सब आज
2068	0	Jitne log aaj shadi me invite hote hai usse jy...	jitne log aaj shadi invite hote hai usse jyada...
2438	Founder at Been There, Done That!	@zyadahorahahai Pati-parmeshwar hota hai! #sar...	zyadahorahahai pati parmeshwar hota hai sarcasm
3171	जिसको भागवा से प्यार नहीं,वो भड़वा मेरा यार नह...	सच मे मुझे बडा दुख होता है 😞\nजब मे किसी विशेष...	सच जब

STEMMING

```
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

ps = PorterStemmer()
```

```
for w in data['Processed Text'][2]:
    print(w, "", ps.stem(w))
```

ए ए
क क
ज ज

ब ब

ब ब
स सऔ औ
र रस स
ब बआ आ
ज ज

data[1:3]

	username	text	Processed Text
1	सत्यमेव जयते\nभगवा ध्वज भारत का ऐतिहासिक एवं...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	एक जब बस और सब आज
2	संस्थापक #भगवारक्षकटीम राम राम दोस्तो 🙏 ...	हमारे गांव में एक मास्टर जी रहते थे, वो मोहल्ल...	एक जब बस और सब आज

data['Processed Text'][1]

'एक जब बस और सब आज'

from nltk import sent_tokenize, word_tokenize

def tokenization_s(data): # same can be achieved for words tokens

s_new = []

for sent in (data['Processed Text'][2]): #For Numpy = sentences[:]

s_token = sent_tokenize(sent)

if s_token != '':

s_new.append(s_token)

return s_new

print(tokenization_s(data))

[['ए'], ['क'], [], ['ज'], ['ब'], [], ['ब'], ['स'], [], ['औ'], ['र'], [], ['स'], ['ब'], [], ['आ'], ['ज']]

from nltk import sent_tokenize, word_tokenize

first_sentence = data['Processed Text'][1]

second_sentence = data['Processed Text'][2]

#split so each word have their own string

first_sentence = first_sentence.split(" ")

second_sentence = second_sentence.split(" ")#join them to remove common duplicate words

```
total= set(first_sentence).union(set(second_sentence))
print(total)
```

```
{'आज', 'सब', 'बस', 'एक', 'और', 'जब'}
```

```
wordDictA = dict.fromkeys(total, 0)
wordDictB = dict.fromkeys(total, 0)
for word in first_sentence:
    wordDictA[word]+=1
```

```
for word in second_sentence:
    wordDictB[word]+=1
```

TF

```
def computeTF(wordDict, doc):
    tfDict = {}
    corpusCount = len(doc)
    for word, count in wordDict.items():
        tfDict[word] = count/float(corpusCount)
    return(tfDict)
```

```
#running our sentences through the tf function:
tfFirst = computeTF(wordDictA, first_sentence)
tfSecond = computeTF(wordDictB, second_sentence)
#Converting to dataframe for visualization
tf = pd.DataFrame([tfFirst, tfSecond])
```

```
print(tf)
```

	आज	सब	बस	एक	और	जब
0	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667
1	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667

```
import math
```

IDF

```
def computeIDF(docList):
    idfDict = {}
    N = len(docList)

    idfDict = dict.fromkeys(docList[0].keys(), 0)
    print(idfDict)
    for word, val in idfDict.items():
```

```

print(word,val)
idfDict[word] = math.log10(N / (float(val) + 1))

return(idfDict)
#inputing our sentences in the log file
ids = computeIDF([wordDictA, wordDictB])
print(ids)

```

```
{'आज': 0, 'सब': 0, 'बस': 0, 'एक': 0, 'और': 0, 'जब': 0}
```

```
आज 0
```

```
सब 0
```

```
बस 0
```

```
एक 0
```

```
और 0
```

```
जब 0
```

```
{'आज': 0.3010299956639812, 'सब': 0.3010299956639812, 'बस': 0.3010299956639812, 'एक': 0.3010299956639812, 'और': 0.3010299956639812, 'जब': 0.3010299956639812}
```

```
tfidf=(tf/ids)
print(tfidf)
```

	आज	सब	बस	एक	और	जब
0	0.553655	0.553655	0.553655	0.553655	0.553655	0.553655
1	0.553655	0.553655	0.553655	0.553655	0.553655	0.553655

```
tf*idf
```

```
tfidfs=(tf*ids)
print(tfidfs)
```

	आज	सब	बस	एक	और	जब
0	0.050172	0.050172	0.050172	0.050172	0.050172	0.050172
1	0.050172	0.050172	0.050172	0.050172	0.050172	0.050172

WORD CLOUD

```

from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
import pandas as pd

```

```

from wordcloud import WordCloud
import matplotlib.pyplot as plt
df_list = []
for i in range(len(data)):
    df_list +=data['text'][i]

```

```
wordcloud = WordCloud(font_path='Lohit-Devanagari.ttf',width = 1000, height = 700,background_color = 'white',min_font_size = 10).generate(data['Processed Text'][1])  
# plot the WordCloud image  
plt.figure(figsize = (18, 8), facecolor = None)  
plt.imshow(wordcloud,interpolation="bilinear")  
plt.axis("off")  
plt.tight_layout(pad = 0)  
plt.show()
```

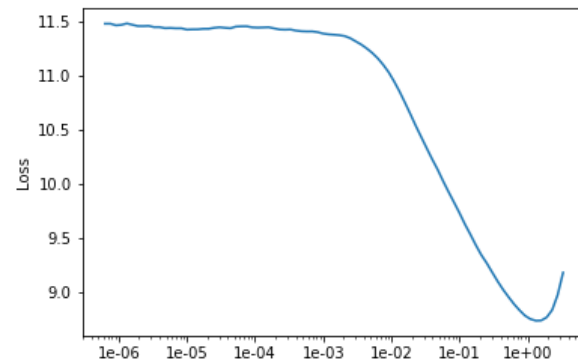


एक बस सब जब और आज

```
learn = language_model_learner(data_lm, AWD_LSTM, pretrained="models/third_hi_lm.pth", drop_mult=0.5)
```

```
learn.lr_find()  
learn.recorder.plot()
```

LR Finder is complete, type {learner_name}.recorder.plot() to see the graph.

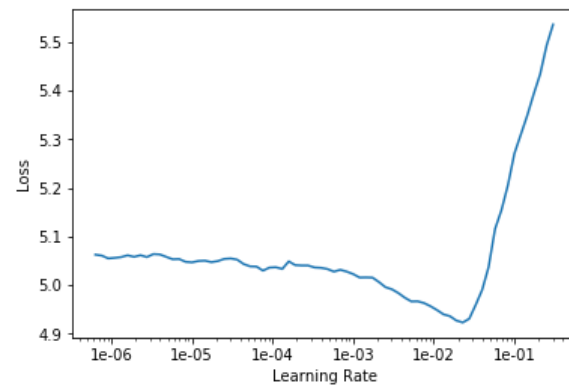


```
learn.fit_one_cycle(1, 3e-02, moms=(0.8, 0.7))
```

epoch	train_loss	valid_loss	accuracy	time
0	6.267576	4.776055	0.280179	00:06

```
learn.unfreeze()
learn.lr_find()
learn.recorder.plot(skip_end=15)
```

LR Finder is complete, type {learner_name}.recorder.plot() to see the graph.



```
learn.fit_one_cycle(15, 3e-04, moms=(0.8, 0.7))
```

epoch	train_loss	valid_loss	accuracy	time
0	5.021370	4.728021	0.285848	00:08
1	4.955410	4.623461	0.296161	00:08
2	4.859788	4.533570	0.303393	00:08
3	4.769974	4.456697	0.305357	00:08
4	4.686345	4.390315	0.314955	00:08
5	4.610135	4.325916	0.321295	00:08
6	4.537043	4.284279	0.324598	00:08
7	4.475051	4.245073	0.329643	00:08
8	4.418847	4.212920	0.331830	00:09
9	4.374950	4.196318	0.331250	00:09
10	4.338069	4.182378	0.332857	00:09
11	4.310486	4.169988	0.332991	00:09

```
TEXT = " में"
N_WORDS = 30
N_SENTENCES = 1
```

```
print("\n".join(learn.predict(TEXT, N_WORDS, temperature=0.75) for _ in range(N_SENTENCES)))
```

में इस जीतना नहीं किया जाएगा कि फिल्म में तरह में पड़ा है । xxbos RT @aahtak : दिल्ली पुलिस में पास 3 साल की इलाके में आजतक पर पुलिस

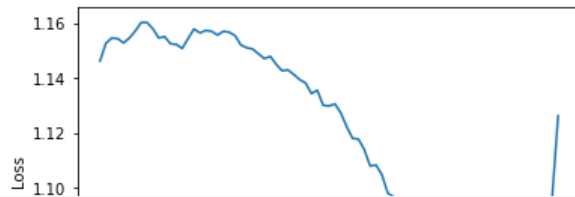
```
learn.export("models/hi_sentence.pkl")
```

```
learn.save_encoder('fine_tuned_enc_9077')
```

```
learn = text_classifier_learner(data_clas, AWD_LSTM, drop_mult=0.5)
learn.load_encoder('fine_tuned_enc_9077')
```

```
learn.lr_find()
learn.recorder.plot()
```

LR Finder is complete, type {learner_name}.recorder.plot() to see the graph.



```
learn.fit_one_cycle(1, 3e-03, moms=(0.8,0.7))
```

epoch	train_loss	valid_loss	accuracy	time
0	0.930926	0.723590	0.643526	00:04

```
learn.freeze_to(-2)
learn.fit_one_cycle(1, slice(1e-03/(2.6**4),1e-03), moms=(0.8,0.7))
```

epoch	train_loss	valid_loss	accuracy	time
0	0.813841	0.671902	0.658953	00:05

```
learn.freeze_to(-3)
learn.fit_one_cycle(3, slice(5e-03/(2.6**4),5e-03), moms=(0.8,0.7))
```

epoch	train_loss	valid_loss	accuracy	time
0	0.853167	0.648381	0.694215	00:07
1	0.741108	0.580622	0.711295	00:07
2	0.720004	0.575250	0.708540	00:07

```
learn.unfreeze()
learn.fit_one_cycle(5, slice(5e-04/(2.6**4),5e-04), moms=(0.8,0.7))
```

epoch	train_loss	valid_loss	accuracy	time
0	0.608293	0.573941	0.720110	00:10
1	0.695433	0.564656	0.728926	00:08
2	0.699171	0.570015	0.727824	00:08
3	0.627100	0.582248	0.731680	00:09
4	0.623950	0.583798	0.731680	00:09


```
learn.fit_one_cycle(5, slice(5e-04/(2.6**4),5e-04), moms=(0.8,0.7))
```

epoch	train_loss	valid_loss	accuracy	time
0	0.622325	0.564834	0.728375	00:09
1	0.619325	0.582253	0.732782	00:09
2	0.644618	0.568589	0.728926	00:08
3	0.627140	0.567003	0.734435	00:09
4	0.561733	0.559429	0.731680	00:10

```
learn.export("classifier_sentiment_hi_v10.pkl")
```

```
learn.predict("आज़तक पर देखिये कितनी स्मार्ट हुई स्मार्ट फेंसिंग Exclusive आज रात 9.30 pm पर @BSF_India")
```

```
(Category positive, tensor(2), tensor([2.6687e-07, 3.2584e-07, 1.0000e+00]))
```

```
learn.predict("चीन के सामने भारत के जांबाज़ सूरमाओं की Exclusive रिपोर्ट कोल्ड डेजर्ट के नाम से मशहूर नेलांग वैली से।... https://t.co/7JBdsL6xaE")
```

```
(Category positive, tensor(2), tensor([3.5505e-06, 2.0953e-06, 9.9999e-01]))
```

```
learn.predict("Female IAS officer molestation दिल्ली: साकेत कोर्ट परिसर में महिला IAS अधिकारी से छेड़छाड़ https://t.co/MtZhnaFYiV")
```

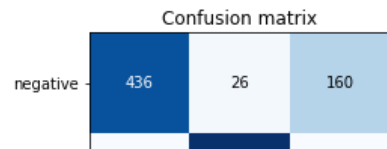
```
(Category neutral, tensor(1), tensor([0.0805, 0.8967, 0.0228]))
```

```
from fastai.vision import *
```

```
preds,y,losses = learn.get_preds(with_loss=True)
```

```
inp = ClassificationInterpretation(learn,preds,y,losses)
```

```
inp.plot_confusion_matrix()
```



```
inp.top_losses(k=1,largest=True)
```

```
torch.return_types.topk(values=tensor([5.4140]), indices=tensor([102]))
```

▼ TESTING OF MODEL

```
test_pos = read_ods("positive.ods",1, headers=False)
```

```
test_pos
```

	column_0	column_1		column_2	column_3		column_4	column_5
0	मुमकिन है कि रिलायंस कैपिटल का शेयर 330 रुपये ...	positive	बुंदेलखंड के किसानों को अबकी बार बड़ी उम्मीदें...	negative	RT @mumbaitak: शिवसेना-बीजेपी गठबंधन में दरार,...		neutral	
1	सितम्बर में इसकी कमिश्निंग संभावित है।	positive	उन्होंने कहा कि चूंकि पूंजी की जरूरत बहुत बड़ी...	negative	RT @mumbaitak: शिवसेना-बीजेपी गठबंधन में दरार,...		neutral	
2	भारतीय कंपनी ने संभावित डील के बारे में कमेंट ...	positive	पश्चिम बंगाल में शासन द्वारा ऐसी शिकायतें लगात...	negative	RT @mumbaitak: बोरवेल में फंसे 6 साल के मासूम ...		neutral	
3	हम अदालत में लड़ाई लड़ने के अलावा इस मामले को ...	positive	शहर में नेटवर्किंग कारोबार का मकड़जाल लगातार फ...	negative	RT @mumbaitak: बीजेपी-शिवसेना गठबंधन पर उद्धव ...		neutral	
4	असल में कोशिश की जा रही है कि संभावित तौर पर क...	positive	हमें नहीं लगता कि आने वाले दिनों में कीमतों मे...	negative	RT @mumbaitak: ठाणे के कोरम मॉल में घुसा तेंदु...		neutral	
5	पोलार्ड और तारे ने असम्भव को सम्भव कर दिखाया था.	positive	मौसम में लगातार परिवर्तन और नलों से आ रहे गंदा...	negative	RT @mumbaitak: ठाणे के कोरम मॉल में तेंदुआ घुस...		neutral	
6	कॉमरेड धनंजय तिवारी की विशेष रूचि और परिश्रम स...	positive	कुशल का मेडिकल इश्यू के कारण कुशल को ये शो छ...	negative	RT @mumbaitak: पुलवामा के शहीदों को श्रद्धांजल...		neutral	
7	क्रिस लिन ने नामुमकिन कैच को अपने गजब के बॉडी ...	positive	एक गोल होने के बाद स्काटलैंड ने भारतीय गोल पर ...	negative	RT @mumbaitak: .@ShivSena -@BJP4India गठबंधन प...		neutral	
8	इससे आपके कंधे ताकतवर बन सकेंगे।	positive	उन्हें धरना देना पड़ा।	negative	RT @aajtak: राम मंदिर पर हम साथ-साथ: उद्धव #Sp...		neutral	
9	वह काफी ताकतवर है और उसके पास विविधता है.	positive	उन्होंने जेपी चौक पर अनिश्चितकालीन धरना आंदोलन ...	negative	@nistula Tyana visarale		neutral	
10	विलेन जितना दमदार और ताकतवर होगा, उतना ही हीरो...	positive	इस बात पर लक्ष्मीनगर की 80 महिलाओं ने शराब भट्...	negative	RT @mumbaitak: खुद की फिल्म पाकिस्तान में रिली...		neutral	
11	लिहाजा इस जीवनदान के बाद मैक्सवेल बल्ले से और ...	positive	इस कारण से किसानों को उनका माल खुले में ही उता...	negative	RT @mumbaitak: CRPF जवान के साथ पुलिस ने की मा...		neutral	
12	उनके ट्विटर में आते ही इस सोशल साइट पर जबरदस्त...	positive	बंगाल को लगातार तीसरी हार का सामना करना पडा और...	negative	RT @mumbaitak: #PulwamaAttack में शहीद जवानों ...		neutral	
13	मोदी के बीसीसीआई से जुड़ते ही उसके रेवेन्यू मे...	positive	पुलिस के लिए सुबूत तलाशना मुश्किल है।	negative	RT @mumbaitak: कपिल शर्मा शो से निकलने के बाद ...		neutral	
14	राजस्थान की जीत में प्रवीण ताम्बे और शेन वॉटसन...	positive	उम्मीदवारों ने हत्या के प्रयास के मुकदमों का श...	negative	RT @mumbaitak: पुलवामा हमले पर क्या बोली महारा...		neutral	
15	मैन आफ द मैच स्मिथ ने कोटला में ऐसी जबरदस्त बल...	positive	परिजनों की तहरीर पर पुलिस ने चार नामजद आरोपियो...	negative	RT @mumbaitak: 'पाकिस्तान मुर्दाबाद' Inयवतमाल ...		neutral	
16	इसमें हीरो हीरोइन दमदार कैमेस्ट्री देखने को मि...	positive	पीलिया संदिग्ध मरीज के खून की जांच में भी वक्त...	negative	RT @mumbaitak: बीजेपी और पीएम मोदी भूल जाएंगे ...		neutral	
17	फिल्म में दमदार भूमिका निभाने के लिए सबसे पहले...	positive	वह उस आदमी का कत्ल कर देती है।	negative	RT @mumbaitak: @PulwamaAttack के विरोध प्रदर्श...		neutral	
18	अभी तक टूर्नामेंट में उनके लिए केवल आरोन फिच ...	positive	उन्होंने शिवांग का कत्ल कर उसके परिजनों से पैस...	negative	RT @mumbaitak: मातोश्री पहुंचे महाराष्ट्र के म...		neutral	
19	नियम उसे खातों की प्राथमिक जांच के आगे जाने के...	positive	इस दौरान रास्ते में घात लगाए बदमाशों ने उनके प...	negative	@ahir_hansraj on #TerroristAttack " hissab chu...		neutral	
20	नियमों के मुताबिक सिनेमा मालिक आपको चश्मा लेने...	positive	अब पूरी दुनिया जान गई है कि मोदी झूठों के बादश...	negative	RT @mumbaitak: महाराष्ट्र में गठबंधन पर सस्पें...		neutral	
21	टाटा मोटर्स का उपयोगी वाहन मोवस पेश.	positive	नगर पंचायत परिसर में जमकर हंगामा हुआ था।	negative	RT @divyeshas: मुंबई ब्लास्ट मामले में भारत को...		neutral	
22	सलामी बल्लेबाज मुरलीधरन गौतम ने भी 30 रन की उप...	positive	दोनों कर्मचारियों पर शराब के नशे में धुत होकर ...	negative	RT @mumbaitak: पैराग्लाइडिंग से सावधाना!! दक्षि...		neutral	
23	कप्तान धौनी और रवींद्र जडेजा ने भी उपयोगी प्रद...	positive	श्वेता की मां अलका मिश्रा को गंभीर चोट आई है।	negative	RT @mumbaitak: 45 वर्षीय के दक्षिण कोरियाई नाग...		neutral	
24	हालांकि प्रयोगात्मक विज्ञान और फिक्शन के बीच ब...	positive	गडकरी ने झारखंड के क्षेत्रीय दलों पर प्रहार किया।	negative	RT @mumbaitak: प्रकाश आंबेडकर की सभा में एक व्...		neutral	
25	हर रैन बसेरे के पास पानी का टैंकर होना अनिवार्...	positive	भाजपा सिर्फ सत्ता की भूखी है।	negative	RT @mumbaitak: मुंबई में CM निवास के पास लगी आ...		neutral	
26	इस मौके पर राकेश कुमार ने कहा कि पढ़ाई के साथ ...	positive	सोमवार की रात्रि डीजल चोरों की ताक में कुछ साह...	negative	RT @rahulkanwal: फैक्ट चेक: नहीं, यह "चायवाला"...		neutral	
27	कांची के पिता वतन की रक्षा करते हुए शहीद हो गए।	positive	यहां भाषण के दौरान लालू ने जब लोगों को डी एरिय...	negative	RT @mumbaitak: अमोल पालेकर विवाद: सरकार के खिल...		neutral	
28	विकसित देशों में ही नहीं बल्कि अपने देश के रोग...	positive	अब बदमाशों के हौसले इतने बुलंद हो गए हैं कि वह...	negative	@lalitncr @mumbaitak @aajtak @mustafashk Kyon ...		neutral	
29	इस सविधा के तहत पीएफ निकासी सहित अन्य भगतान	positive	रुह के खिलाफ जआ एकट में कार्यवाई की गई।	negative	Is this #TwitterDefiesIndia started because		neutral	

...	...	positive	...	negative	...	neutral
...
655	भारत की उन्नति भारतीयों पर निर्भर है	positive	पुलिस अभी इससे बेखबर है पर सूत्रों का कहना है ...	negative	RT @aajtak: 'हिंदू आतंक' से कांग्रेस ने बदनाम ...	neutral
656	राम ने अपने आप को निर्दोष सिद्ध करने के लिए बह...	positive	रिलायंस कंपनी के कारण उनका आर्थिक और मानसिक उत...	negative	RT @aajtak: क्या हिंदू आतंक पर कांग्रेस को घेर...	neutral
657	उसने अपने घर की दीवारों पर बहुत सुंदर चित्रांक...	positive	मानसिक तनाव से भी अस्थमा के अटैक होने की संभाव...	negative	RT @aajtak: हिंदुओं को कांग्रेस से डरा रहे हैं...	neutral
658	याम चित्रकला प्रतियोगिता में प्रथम आया	positive	उनका मानना है कि ये एक ऐसा माध्यम है, जो सीधे ...	negative	हिंदू का सगा कौन, हिंदू को ठगा कौन?\nअसली हिंद...	neutral
659	नाटक का प्रस्तुतिकरण अच्छा था	positive	देश के लोगों को अपनी मानसिकता बदलने की जरूरत है।	negative	चुनाव के घमासान में आपकी बात, आपके साथ.\nअब से...	neutral
660	ताजमहल कुशल शिल्पियों की एक अनुपम कृति है	positive	अभिभावक आरपार की लड़ाई लड़ने के मूड में हैं।	negative	RT @uttamkrbhardwaj: मेहनत, हिम्मत और लगन से क...	neutral
661	आच्छाद से वस्तुएँ सुरक्षित रहती हैं	positive	सूत्रों का कहना है कि वह पार्टी में बिल्कुल जश...	negative	फिर अल्पसंख्यक वोट के भरोसे कांग्रेस? https://...	neutral
662	भगवान कृष्ण विदुर के अतिथि-सत्कार से बहुत ही प...	positive	इस बार युवराज काफी सीरियस मूड में हैं।	negative	चारे में भी बँटवारा कर लोगे, \nचिटफ़ंड भी हज़म...	neutral
663	विदेशी लोग भारतीय शिष्टाचार से बहुत प्रभावित ह...	positive	गंभीर अवस्था में कई बच्चों को भर्ती भी करना पड़...	negative	RT @aajtak: राहुल गांधी के वायनाड से चुनाव लड़...	neutral
664	सृजनशील गृहणी घर की शोभा बढ़ाती है।	positive	दरवाजा खोलकर अंदर देखा तो युवक झुलसी अवस्था मे...	negative	RT @aajtak: एक से भले दो..!\nदेखिए #Dangal @c...	neutral
665	इस कुर्सी की नक्काशी बहुत सुन्दर है	positive	घायल अवस्था में पुलिस व्यापारी को अस्पताल ले ग...	negative	पप्पू जी का लाल टी-शर्ट नहीं देखे मिश्रा जी? 🙄...	neutral
666	यह मूर्तिकार बहुत अच्छी नक्काशी करता है	positive	इसी दौरान पीछे से आए व्यापारियों ने रविंद्र को...	negative	जितनी बार बगदादी नहीं मरा, उससे ज़्यादा बार तो...	neutral
667	दूल्हे के सिर पर मौर सुशोभित था	positive	मैंने टीम के बीच सौहार्द्र की भावना को सर्वाधि...	negative	बस ये बात हवाओं को बताए रखना,\nरौशनी होगी चिरा...	neutral

```

result = []
count_pos=0
count_pos_neg=0
count_neg=0
count_neg_pos=0
count_neu=0
count_neu_pos=0
for text in test_pos.column_0:
    #print(str(learn.predict(str(text))[0]))
    if str(learn.predict(str(text))[0]) == "positive":
        count_pos+=1
    if str(learn.predict(str(text))[0]) == "negative":
        count_pos_neg+=1

for text in test_pos.column_2:
    #print(str(learn.predict(str(text))[0]))
    if str(learn.predict(str(text))[0]) == "negative":
        count_neg+=1
    if str(learn.predict(str(text))[0]) == "positive":
        count_neg_pos+=1

for text in test_pos.column_4:
    #print(str(learn.predict(str(text))[0]))
    if str(learn.predict(str(text))[0]) == "neutral":
        count_neu+=1

```

```
if str(learn.predict(str(text))[0]) == "positive":  
    count_neu_pos+=1
```

Positive Set

```
count_pos/686 , count_pos_neg/686 , 1-count_pos/686-count_pos_neg/686  
  
(0.5539358600583091, 0.4446064139941691, 0.001457725947521804)
```

Negative Set

```
count_neg/686 , count_neg_pos/686 , 1-count_neg/686-count_neg_pos/686  
  
(0.793002915451895, 0.2055393586005831, 0.0014577259475218596)
```

Neutral Set

```
count_neu/686 , count_neu_pos/686 , 1-count_neu/686-count_neu_pos/686  
  
(0.8221574344023324, 0.16034985422740525, 0.017492711370262398)
```