

✳ Text Summarization using NLP techniques

[Word Frequency, TF-IDF, and BART]

🏆 Sunil Chhetri to be felicitated for 150th international before FIFA World Cup 2026 qualifier

Sunil Chhetri first donned the Senior National Team jersey on June 12, 2005, in a friendly match against Pakistan in Quetta and has went on to represent the tricolour 148 more times.

Published : Mar 23, 2024 18:55 IST, Chennai



Chhetri is the third-highest active goalscorer (93 goals) in international football, just below Cristiano Ronaldo (128 goals) and Lionel Messi (106 goals). | Photo Credit: PTI

Article → <https://sportstar.thehindu.com/football/indian-football/sunil-chhetri-felicitated-150-international-appearances-fifa-world-cup-qualifier-india-vs-afghanistan/article67984404.ece>

```
## input text article
```

```
article_text='''Sunil Chhetri, the captain of the Indian Men's National team, will be felicitated by the All India Football F  
who is expected to play his 150th Senior International match in the FIFA World Cup 2026 qualifier on March 26.  
India, after drawing with Afghanistan in the away leg, will host it at the Indira Gandhi International Stadium in Guwahati.  
This round will also serve as the second round of preliminary joint qualification for the AFC Asian Cup 2027.  
"It's been an awe-inspiring and staggering cruise that all of us have been fortunate to witness since 2005. Chhetri playing h  
international match is an extraordinary accomplishment that will go a long way in keeping the Indian football flag flying hig  
the president of AIFF, said.
```

```
Chhetri first donned the Senior National Team jersey on June 12, 2005, in a friendly match against Pakistan in Quetta. He als  
Since then, he has made 149 appearances for the National Team, netting a record 93 goals. Chhetri holds the unique record of  
goal in his first, 25th, 50th, 75th, 100th and 125th match for the Blue Tigers.  
He is also the third-highest active goalscorer (93 goals) in international football, just below Cristiano Ronaldo (128 goals)  
"In many football fans' eyes, Chhetri is almost the synonym of the Indian Men's National Team, something that is not far from  
He is a fabulous footballer, a cool captain, and a star striker," M Satyanarayan, the Acting General Secretary of AIFF, said.
```

✓ Import Modules

```
!pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.4.0)
Requirement already satisfied: regex<=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.2)
```

```
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer
```

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

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

▼ Data Preprocessing

```
#converting the article text into lower case
article_text = article_text.lower()
article_text
```

```
'sunil chhetri, the captain of the indian men's national team, will be felici
tated by the all india football federation (aiff),\nwho is expected to play h
is 150th senior international match in the fifa world cup 2026 qualifier on m
arch 26.\nindia, after drawing with afghanistan in the away leg, will host it
at the indira gandhi international stadium in guwahati.\nthis round will also
serve as the second round of preliminary joint qualification for the afc asia
n cup 2027.\n"it's been an awe-inspiring and staggering cruise that all of us
have been fortunate to witness since 2005. chhetri playing his 150th\ninternat
ional match is an extraordinary accomplishment that will go a long way in ke
```

► Tokenization

```
# split into sentence list
sentence_list = sent_tokenize(article_text)
sentence_list
```

```
['sunil chhetri, the captain of the indian men's national team, will be felicitated by the all india football federation (aiff),\nwho is expected to play his 150th senior international match in the fifa world cup 2026 qualifier on march 26.\nindia, after drawing with afghanistan in the away leg, will host it at the indira gandhi international stadium in guwahati.',
 'this round will also serve as the second round of preliminary joint qualification for the afc asian cup 2027.',
 '"it's been an awe-inspiring and staggering cruise that all of us have been fortunate to witness since 2005. chhetri playing his 150th\ninternational match is an extraordinary accomplishment that will go a long way in keeping the indian football flag flying high," kalyan chaubey,\nthe president of aiff, said.',
 'chhetri first donned the senior national team jersey on june 12, 2005, in a friendly match against pakistan in quetta.',
 'he also scored in the 1-1 draw.',
 'since then, he has made 149 appearances for the national team, netting a record 93 goals.',
 'chhetri holds the unique record of scoring at least one\ngoal in his first, 25th, 50th, 75th, 100th and 125th match for the blue tigers.',
 'he is also the third-highest active goalscorer (93 goals) in international football, just below cristiano ronaldo (128 goals) and lionel messi (106 goals).',
 '"in many football fans' eyes, chhetri is almost the synonym of the indian men's national team, something that is not far from the truth.',
 'he is a fabulous footballer, a cool captain, and a star striker," m satyanarayan, the acting general secretary of aiff, said.']
```

```
# remove spaces, punctuations and numbers
clean_text = re.sub('[^a-zA-Z]', ' ', article_text)
clean_text = re.sub('\s+', ' ', clean_text)
clean_text
```

```
'sunil chhetri the captain of the indian men s national team will be felicita
ted by the all india football federation aiff who is expected to play his th
senior international match in the fifa world cup qualifier on march india aft
er drawing with afghanistan in the away leg will host it at the indira gandhi
international stadium in guwahati this round will also serve as the second ro
und of preliminary joint qualification for the afc asian cup it s been an awe
inspiring and staggering cruise that all of us have been fortunate to witness
since chhetri playing his th international match is an extraordinary accompli
shment that will go a long way in keeping the indian football flag flying hig
```

► Removal of stop words

```
# removing stop words from the clean text.
stopwords = nltk.corpus.stopwords.words('english')
stopwords[0:10]

↩ ['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're"]
```

✓ 1.Text Summarization Using Word Frequency

► Word Frequencies

```
word_frequencies = {}
for word in nltk.word_tokenize(clean_text):
    if word not in stopwords:
        if word not in word_frequencies:
            word_frequencies[word] = 1
        else:
            word_frequencies[word] += 1
```

```
word_frequencies
```

```
↩ {'sunil': 1,
  'chhetri': 5,
  'captain': 2,
  'indian': 3,
  'men': 2,
  'national': 4,
  'team': 4,
  'felicitated': 1,
  'india': 2,
  'football': 4,
  'federation': 1,
  'aiff': 3,
  'expected': 1,
  'play': 1,
  'th': 7,
  'senior': 2,
  'international': 4,
  'match': 4,
  'fifa': 1,
  'world': 1,
  'cup': 2,
  'qualifier': 1,
  'march': 1,
  'drawing': 1,
  'afghanistan': 1,
  'away': 1,
  'leg': 1,
  'host': 1,
  'indira': 1,
  'gandhi': 1,
  'stadium': 1,
  'guwahati': 1,
  'round': 2,
  'also': 3,
  'serve': 1,
  'second': 1,
  'preliminary': 1,
  'joint': 1,
  'qualification': 1,
  'afc': 1,
  'asian': 1,
  'awe': 1,
  'inspiring': 1,
  'staggering': 1,
  'cruise': 1,
  'us': 1,
  'fortunate': 1,
  'witness': 1,
  'since': 2,
  'playing': 1,
  'extraordinary': 1,
  'accomplishment': 1,
  'go': 1,
  'long': 1,
  'way': 1,
  'keeping': 1,
  'flag': 1,
  'flying': 1,
  ...}
```

► Calculate Sentence Scores

word_frequencies

sentence_scores

4/13

```
'since then, he has made 149 appearances for the national team, netting a record 93 goals.': 2.7142857142857144,
'chhetri holds the unique record of scoring at least one\ngoal in his first, 25th, 50th, 75th, 100th and 125th match
for the blue tigers.': 2.9999999999999999,
'he is also the third-highest active goalscorer (93 goals) in international football, just below cristiano ronaldo
(128 goals) and lionel messi (106 goals).': 4.142857142857142,
'"in many football fans' eyes, chhetri is almost the synonym of the indian men's national team, something that is not
far from the truth.': 4.285714285714285,
'he is a fabulous footballer, a cool captain, and a star striker," m satyanarayan, the acting general secretary of
aiff, said.': 2.285714285714285}
```

► Summary

```
# get top 2 sentences
import heapq
summary_1 = heapq.nlargest(2, sentence_scores, key=sentence_scores.get)
```

```
# Print summary
print("Summary:")
print(summary_1)
```

```
Summary:
["in many football fans' eyes, chhetri is almost the synonym of the indian men's national team, something that is not f
```

```
["in many football fans' eyes, chhetri is almost the synonym of the indian men's national team, something that is not far
from the truth.', 'he is also the third-highest active goalscorer (93 goals) in international football, just below cristiano
ronaldo (128 goals) and lionel messi (106 goals).']
```

✓ 2.Text Summarization Using TF-IDF

```
# Tokenize the text into words and remove stopwords
stop_words = set(stopwords)
word_tokens = [word_tokenize(sentence.lower()) for sentence in sentence_list]
filtered_words = [[word for word in words if word.isalnum() and word not in stop_words] for words in word_tokens]
```

```
# Convert list of words back to sentences
preprocessed_sentences = [' '.join(words) for words in filtered_words]
```

► TF-IDF vectorizer

```
# Create TF-IDF vectorizer
tfidf_vectorizer = TfidfVectorizer()
```

```
# Fit and transform the preprocessed sentences
tfidf_matrix = tfidf_vectorizer.fit_transform(preprocessed_sentences)
```

This line fits the TF-IDF vectorizer to the preprocessed sentences and transforms them into a TF-IDF weighted document-term matrix. Each row in `tfidf_matrix` represents a sentence, and each column represents a word in the entire corpus of sentences. The values in the matrix represent the TF-IDF score of each word in each sentence.

```
# Calculate TF-IDF scores for each word
feature_names = tfidf_vectorizer.get_feature_names_out()
feature_names
```

```
array(['100th', '106', '12', '125th', '128', '149', '150th', '2005',
      '2026', '2027', '25th', '50th', '75th', '93', 'accomplishment',
      'acting', 'active', 'afc', 'afghanistan', 'aiff', 'almost', 'also',
      'appearances', 'asian', 'away', 'blue', 'captain', 'chaubey',
      'chhetri', 'cool', 'cristiano', 'cruise', 'cup', 'donned', 'draw',
      'drawing', 'expected', 'extraordinary', 'eyes', 'fabulous', 'fans',
      'far', 'federation', 'felicitated', 'fifa', 'first', 'flag',
      'flying', 'football', 'footballer', 'fortunate', 'friendly',
      'gandhi', 'general', 'go', 'goal', 'goals', 'goalscorer',
      'guwahati', 'high', 'holds', 'host', 'india', 'indian', 'indira',
      'international', 'jersey', 'joint', 'june', 'kalyan', 'keeping',
      'least', 'leg', 'lionel', 'long', 'made', 'many', 'march', 'match',
      'men', 'messi', 'national', 'netting', 'one', 'pakistan', 'play',
      'playing', 'preliminary', 'president', 'qualification',
      'qualifier', 'quetta', 'record', 'ronaldo', 'round', 'said',
      'satyanarayan', 'scored', 'scoring', 'second', 'secretary',
      'senior', 'serve', 'since', 'something', 'stadium', 'staggering',
```