

Word2Vec

```
import nltk
```

```
from gensim.models import Word2Vec
```

```
from nltk.corpus import stopwords
```

```
import re
```

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

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

```
paragraph = ""I have three visions for India. In 3000 years of our history, people from all over
```

```
the world have come and invaded us, captured our lands, conquered our minds.
```

```
From Alexander onwards, the Greeks, the Turks, the Moguls, the Portuguese, the British,
```

```
the French, the Dutch, all of them came and looted us, took over what was ours.
```

```
Yet we have not done this to any other nation. We have not conquered anyone.
```

```
We have not grabbed their land, their culture,
```

```
their history and tried to enforce our way of life on them.
```

```
Why? Because we respect the freedom of others. That is why my
```

```
first vision is that of freedom. I believe that India got its first vision of
```

```
this in 1857, when we started the War of Independence. It is this freedom that
```

```
we must protect and nurture and build on. If we are not free, no one will respect us.
```

```
My second vision for India's development. For fifty years we have been a developing nation.
```

```
It is time we see ourselves as a developed nation. We are among the top 5 nations of the world
```

```
in terms of GDP. We have a 10 percent growth rate in most areas. Our poverty levels are falling.
```

Our achievements are being globally recognised today. Yet we lack the self-confidence to

see ourselves as a developed nation, self-reliant and self-assured. Isn't this incorrect?

I have a third vision. India must stand up to the world. Because I believe that unless India

stands up to the world, no one will respect us. Only strength respects strength. We must be

strong not only as a military power but also as an economic power. Both must go hand-in-hand.

My good fortune was to have worked with three great minds. Dr. Vikram Sarabhai of the Dept. of

space, Professor Satish Dhawan, who succeeded him and Dr. Brahm Prakash, father of nuclear material.

I was lucky to have worked with all three of them closely and consider this the great opportunity of my life.

I see four milestones in my career. india is great compare to other country "" #
your full text here

Preprocessing

```
text = re.sub(r'\[[0-9]*\]', ' ', paragraph)
```

```
text = re.sub(r'\s+', ' ', text)
```

```
text = text.lower()
```

```
text = re.sub(r'\d', ' ', text)
```

```
text = re.sub(r'\s+', ' ', text)
```

Tokenization

```
sentences = nltk.sent_tokenize(text)
```

```
sentences = [nltk.word_tokenize(sentence) for sentence in sentences]
```

Remove stopwords

```

stop_words = set(stopwords.words('english'))

for i in range(len(sentences)):
    sentences[i] = [word for word in sentences[i] if word not in stop_words]

# Train Word2Vec
model = Word2Vec(sentences, vector_size=100, window=5, min_count=1)

# Vocabulary
words = list(model.wv.index_to_key)
print("Vocabulary:", words)

# Word vector example
if 'war' in model.wv.key_to_index:
    print("\nVector for 'war':", model.wv['war'])

# Most similar words
for target in ['freedom', 'vikram']:
    if target in model.wv.key_to_index:
        print(f"\nMost similar to '{target}':", model.wv.most_similar(target))
    else:
        print(f"\n'{target}' not in vocabulary")

```