

NLP ASSESSMENT

1. Copy the next paragraph and answer the questions that follow:
 - a) Word_tokenize and sent_tokenize
 - b) Using stop words eliminate most common words, do stemming and lemmatization.

For decades the All-India Congress under the leadership of Mohandas K. Gandhi struggled to rally the millions of British-ruled peoples in the Indian subcontinent. Like similar movements in other countries, it early felt the need for a distinctive symbol that could represent its nationalist objectives. In 1921 a university lecturer named Pingali (or Pinglay) Venkayya presented a flag design to Gandhi that consisted of the colours associated with the two principal religions, red for the Hindus and green for the Muslims. To the centre of the horizontally divided flag, Lala Hans Raj Sondhi suggested the addition of the traditional spinning wheel, which was associated with Gandhi's crusade to make Indians self-reliant by fabricating their own clothing from local fibres.

Gandhi modified the flag by adding a white stripe in the centre for the other religious communities in India, thus also providing a clearly visible background for the spinning wheel. In May 1923 at Nagpur, during peaceful protests against British rule, the flag was carried by thousands of people, hundreds of whom were arrested. The Congress flag came to be associated with nationhood for India, and it was officially recognized at the annual meeting of the party in August 1931. At the same time, the current arrangement of stripes and the use of deep saffron instead of red were approved. To avoid the sectarian associations of the original proposal, new attributions were associated with the saffron, white, and green stripes. They were said to stand for, respectively, courage and sacrifice, peace and truth, and faith and chivalry. During World War II Subhas Chandra Bose used this flag (without the spinning wheel) in territories his Japanese-aided army had captured.

2. Copy the paragraph and apply the bag-of-words approach; Also, identify the bag-of-vector for each sentence.

Construction of the mausoleum was essentially completed in 1643. but work continued on other phases of the project for another 10 years. The Taj Mahal complex is believed to have been completed in its entirety in 1653 at a cost estimated at the time to be around 32 million. The construction project employed some 20,000 artisans under the guidance of a board of architects led by the court architect to the emperor, Ustad Ahmad Lahauri. Various types of symbolism have been employed in the Taj to reflect natural beauty and divinity.

3. Copy the paragraph and apply TFIDF method and find the feature vector for each sentence.

Referred to as the Venice of the East, Alappuzha has always enjoyed an important place in the maritime history of Kerala. Today, it is famous for its boat races, backwater holidays, beaches, marine products and coir industry. Alappuzha Beach is a popular picnic spot.