1. Write a Python code using NLP to Pre-Process the text data and convert Text-Numeric vectors.
   I. Use Tokenization, Stopword removal, Stemming/Lemmatization , text preprocess logic using NLTK
   II. Use SKLearn for converting Text-Numeric vectors using TF-IDF model

consider novel.txt as text document for implementing question 1.

Ans:

```
 import nltk

from nltk.corpus import stopwords

from nltk.stem import WordNetLemmatizer

from nltk.tokenize import word_tokenize

from sklearn.feature_extraction.text import TfidfVectorizer

# Step 1: Load NLTK resources

nltk.download('punkt')

nltk.download('stopwords')

nltk.download('wordnet')

# Step 2: Read the text document

with open('novel.txt', 'r', encoding='utf-8') as file:

    text = file.read()

# Step 3: Tokenization

tokens = word_tokenize(text)

# Step 4: Stopword Removal

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

filtered_tokens = [token for token in tokens if token.lower() not in stopwords]

# Step 5: Stemming/Lemmatization

lemmatizer = WordNetLemmatizer()

lemmatized_tokens = [lemmatizer.lemmatize(token) for token in filtered_tokens]

# Step 6: Preprocessed text

preprocessed_text = ' '.join(lemmatized_tokens)
```

```
# Step 7: TF-IDF vectorization

vectorizer = TfidfVectorizer()

text_vectors = vectorizer.fit_transform([preprocessed_text])

text_numeric_vectors = text_vectors.toarray()


# Step 8: Print the text-numeric vectors

print(text_numeric_vectors)
```