

**LAB FILE**

**Of**

**Natural Language Processing**

# Submitted by:

## Piyush Upadhyay

## 2115800020 (18)

# Submitted To:

# Dr. Gourav Bathla

**Department of Computer Engineering & Applications**

**Institute of Engineering & Technology**

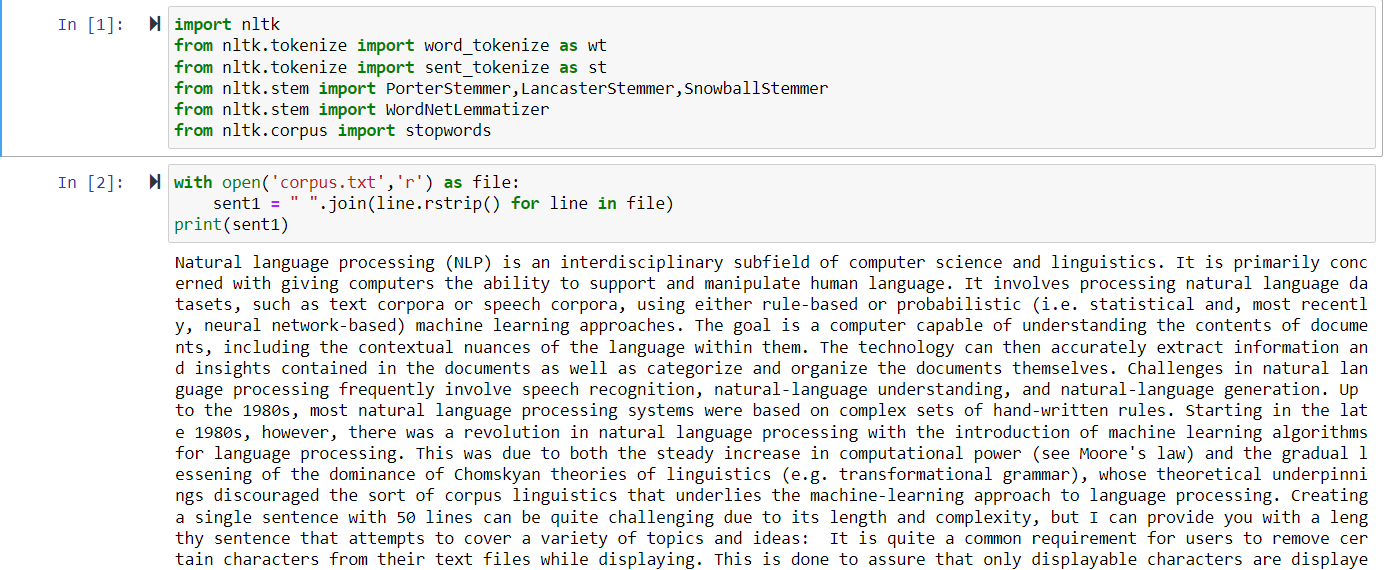
# GLA University Mathura- 281406, INDIA

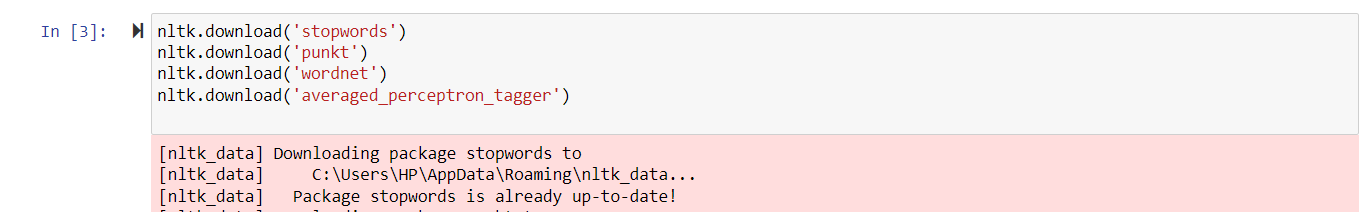
**2024**

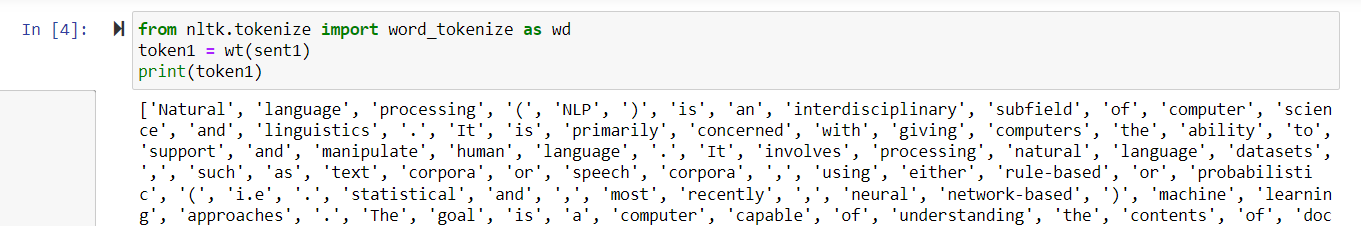
# Table Of Contents

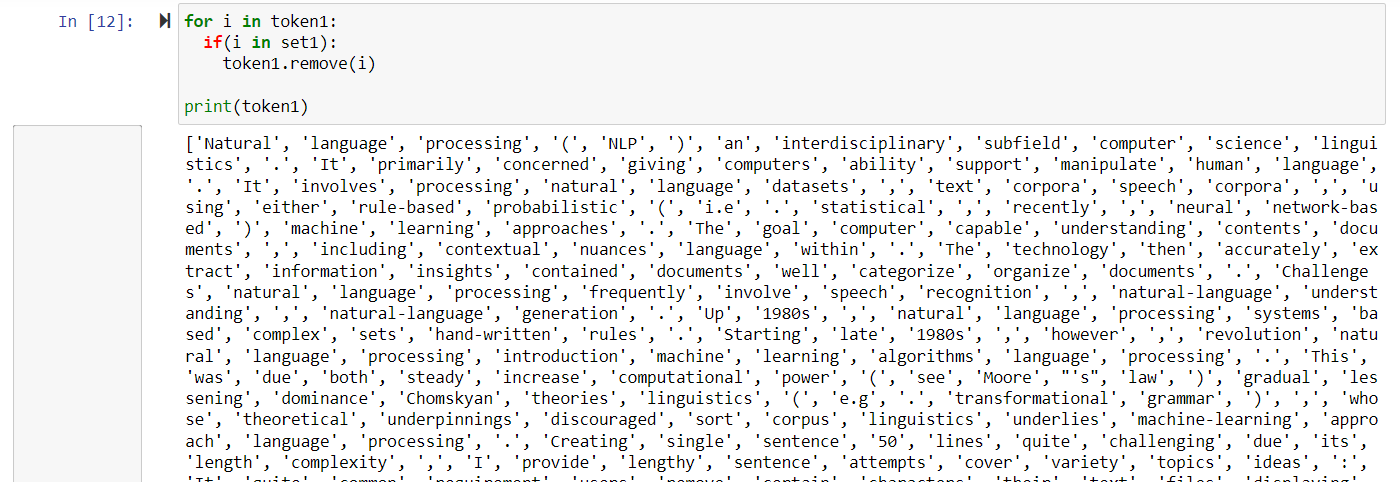
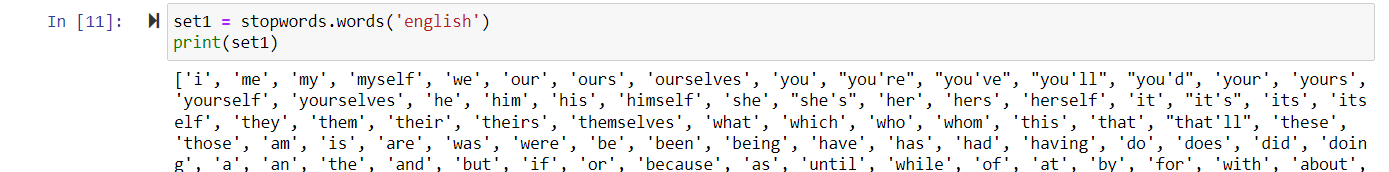
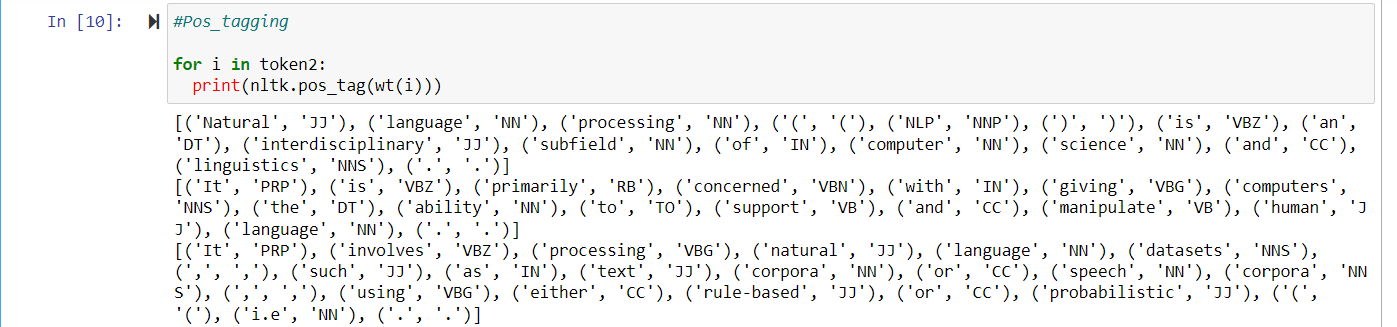
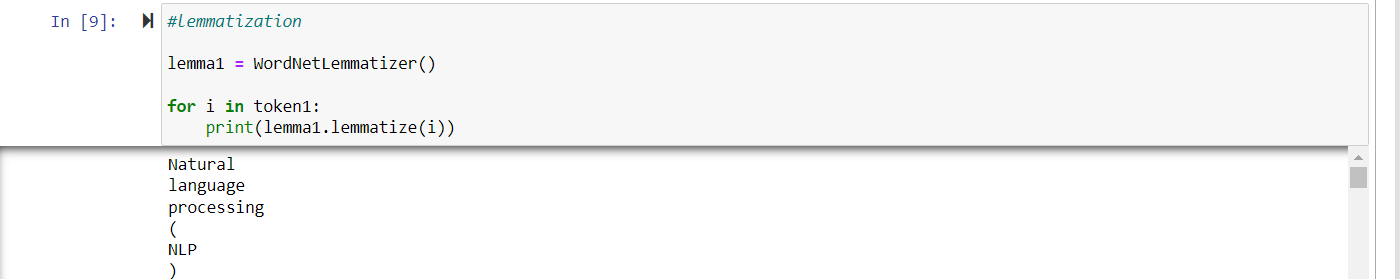
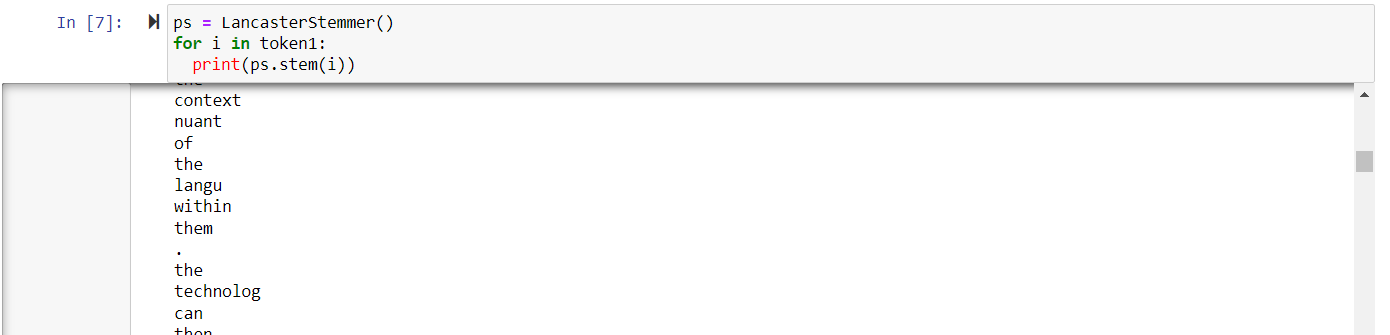
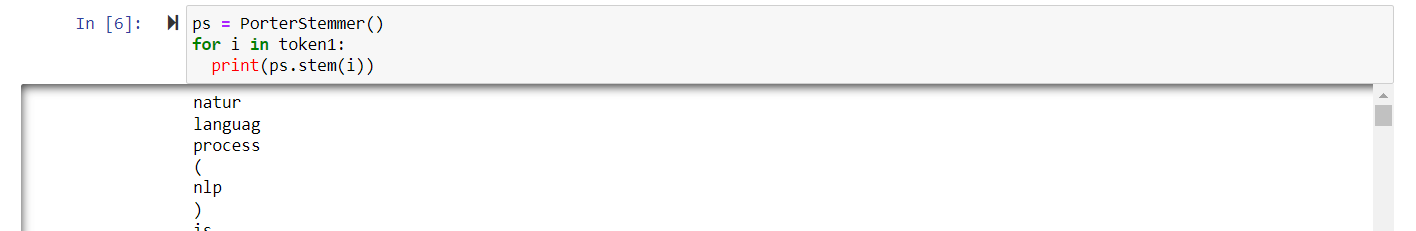
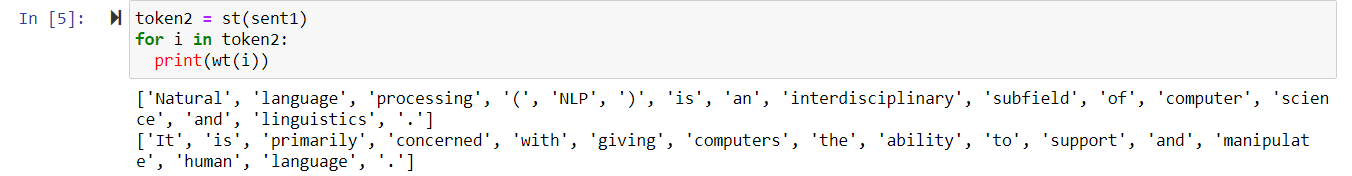
|  |  |  |
| --- | --- | --- |
| **Topics** | **Page Number** | |
| 1. **Implementation of tokenization, stemming, lemmatization, stop-words removal, POS tagging** | | **3** |
| 1. **Implementation of text to vectors using CountVectorizer** | |  |
| 1. **Implementation of regular expression** | |  |
| 1. **Implementation of corpus- Brown, Gutenberg, Indian, Treebank, Wordnet** | |  |
| 1. **Implementation of WordNet** | |  |
| 1. **Implementation of TF/TDF using manual computation and using TFldfvectorizer** | |  |
| 1. **Word Embedding: Bag-of-Words, TF/IDF** | |  |
| 1. **Word Embedding: Glove, BERT, Word2Vec** | |  |
| 1. **Implementation of Spam/Not Spam email using TFIdfvectorizer and Countvectorizer** | |  |
| 1. **Sentiment Analysis** | |  |
| 1. **Duplicate Question Detection using BERT** | |  |
| 1. **Recommendation System** | |  |

1. **Implementation of tokenization, stemming, lemmatization, stop-words removal, POS tagging**

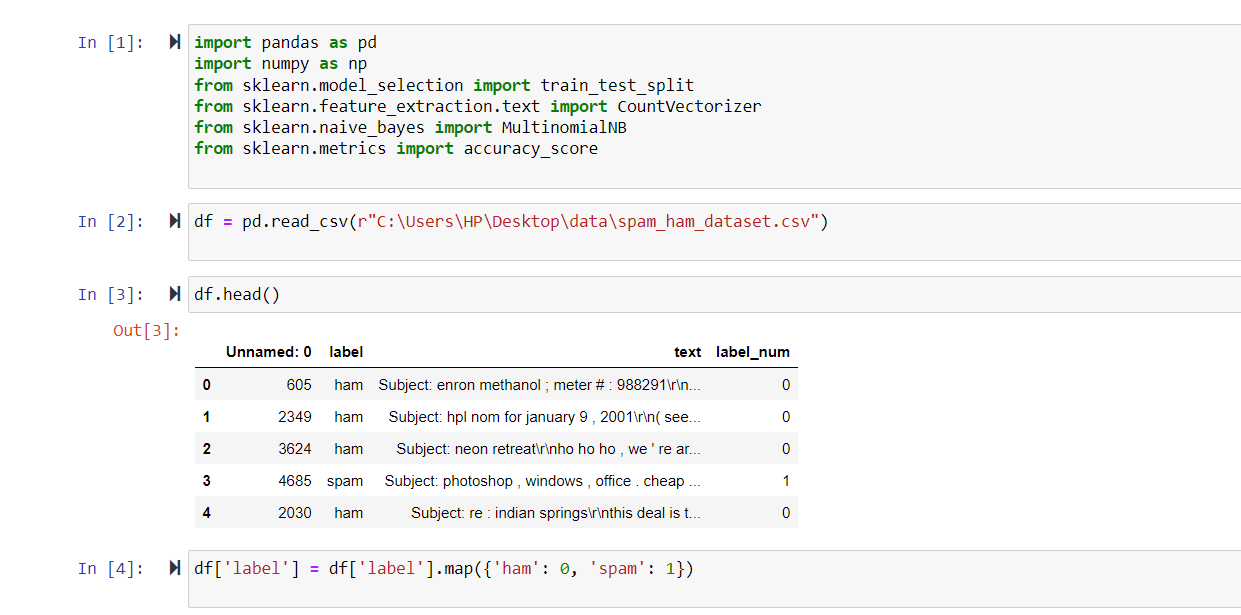
****

****

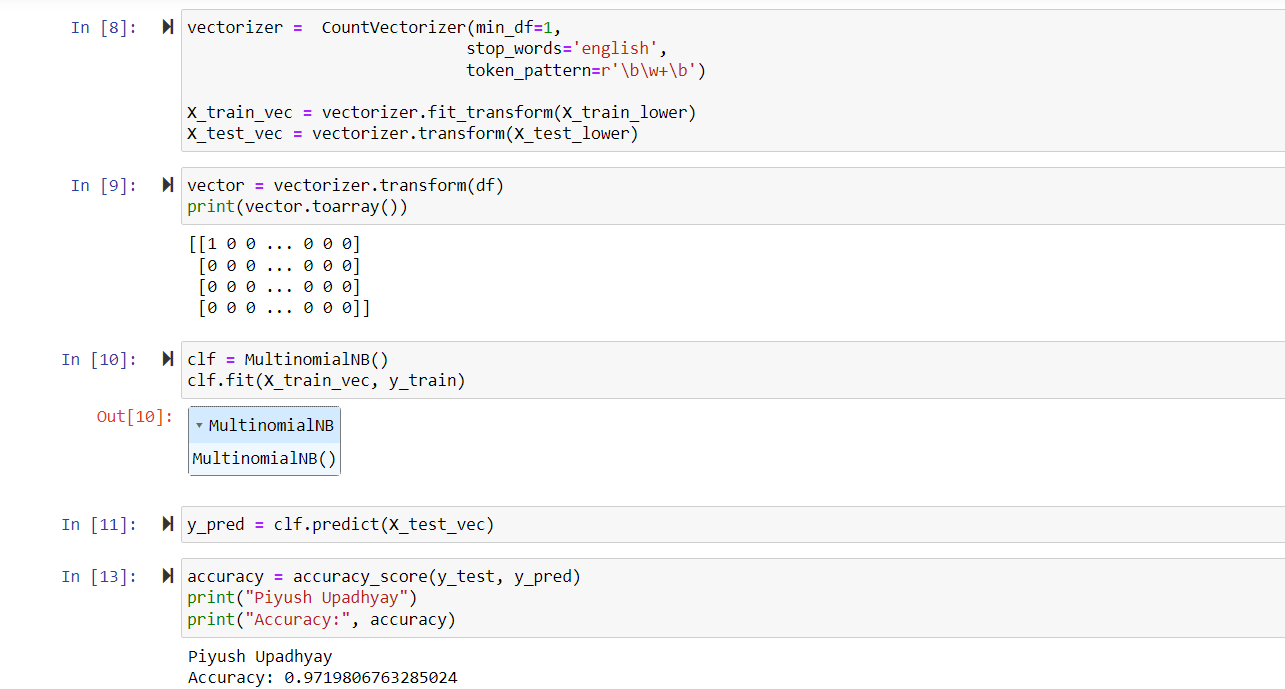
****

****

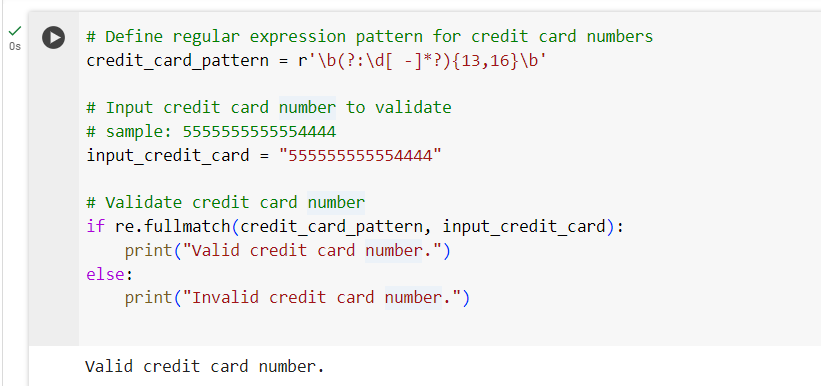
1. **Implementation of text to vectors using CountVectorizer**

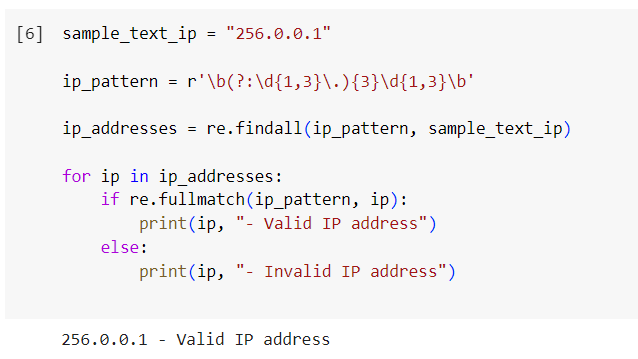
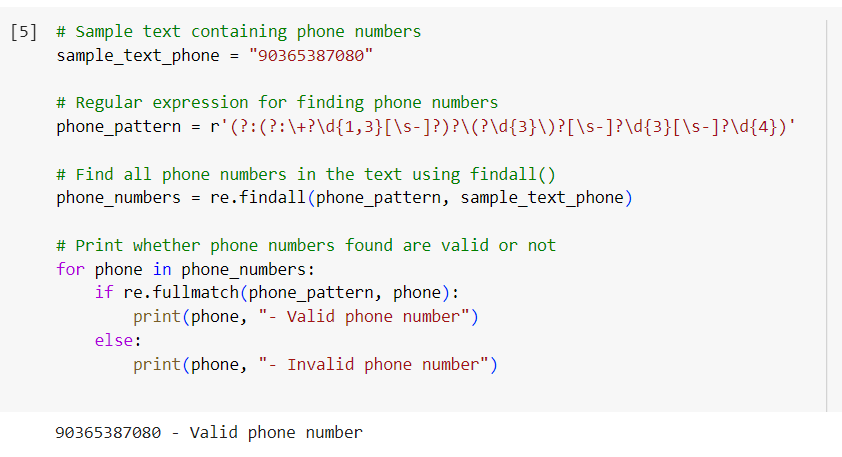
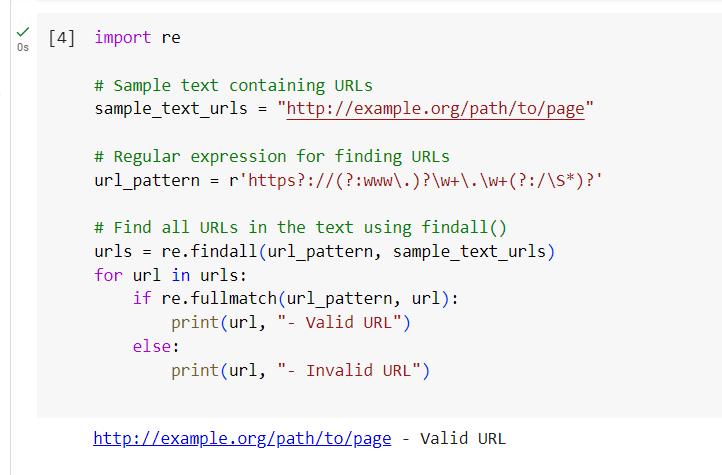




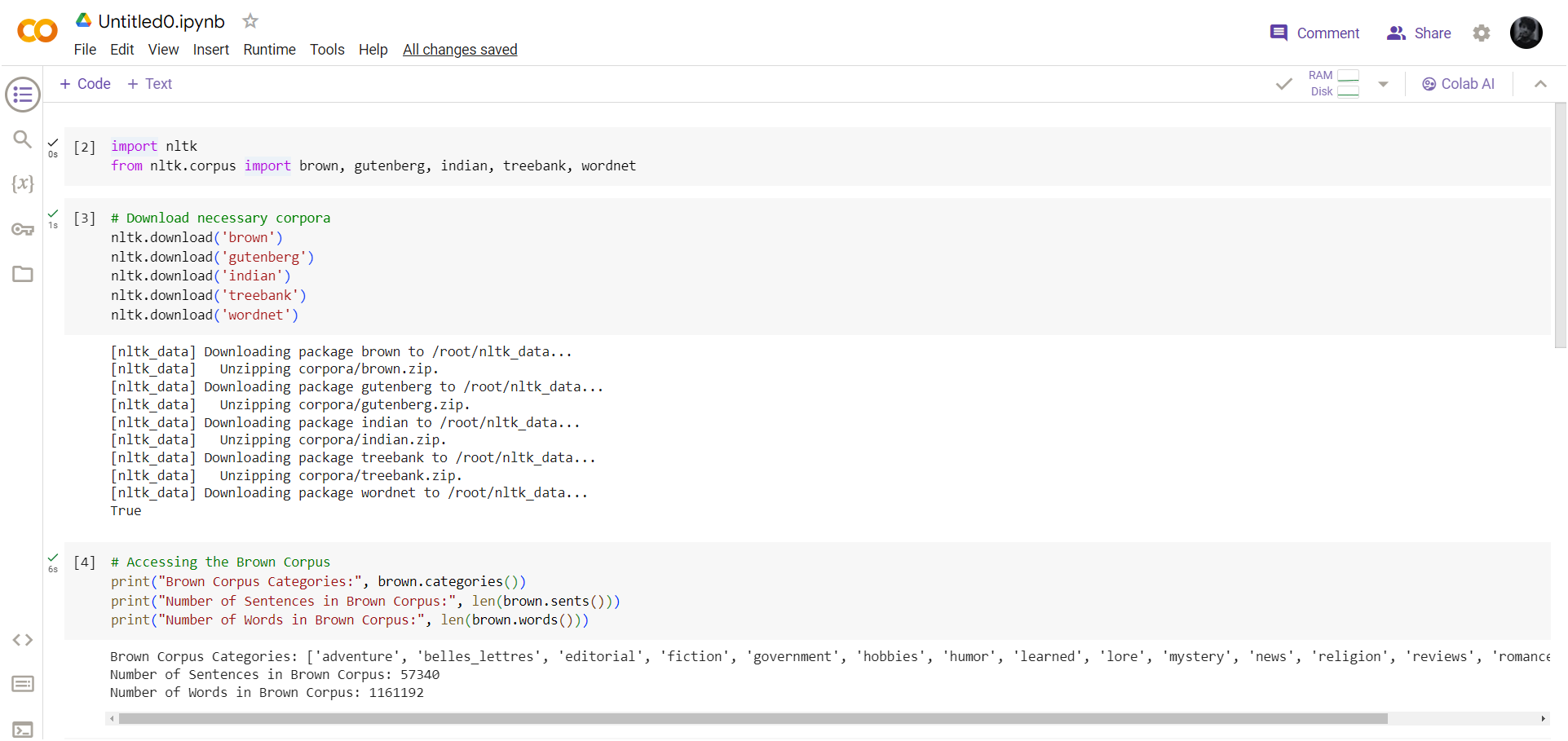


1. **Implementation of regular expression**

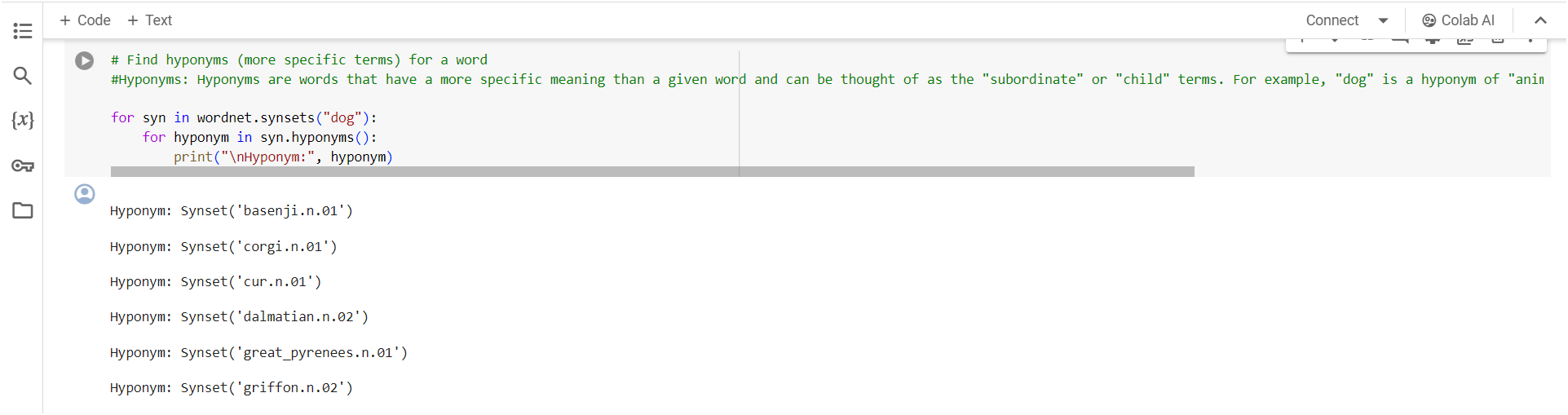
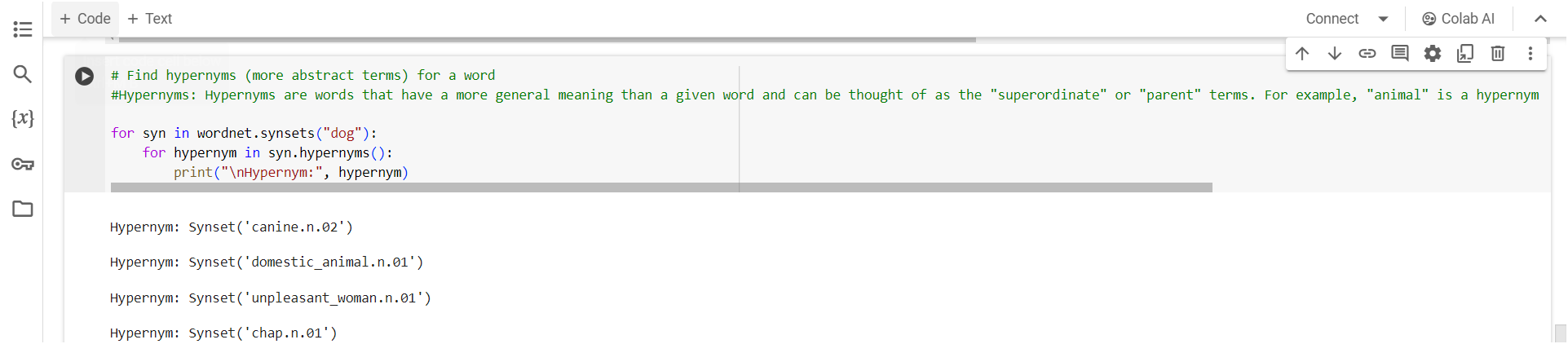




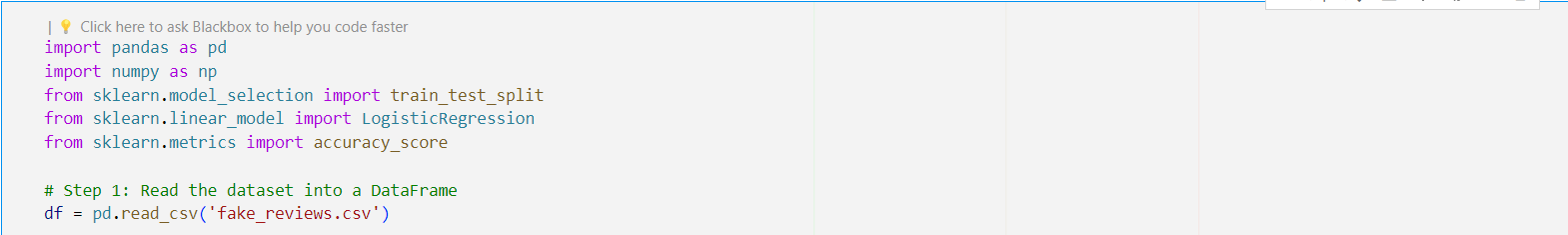
**4. Implementation of corpus- Brown, Gutenberg, Indian, Treebank, Wordnet**

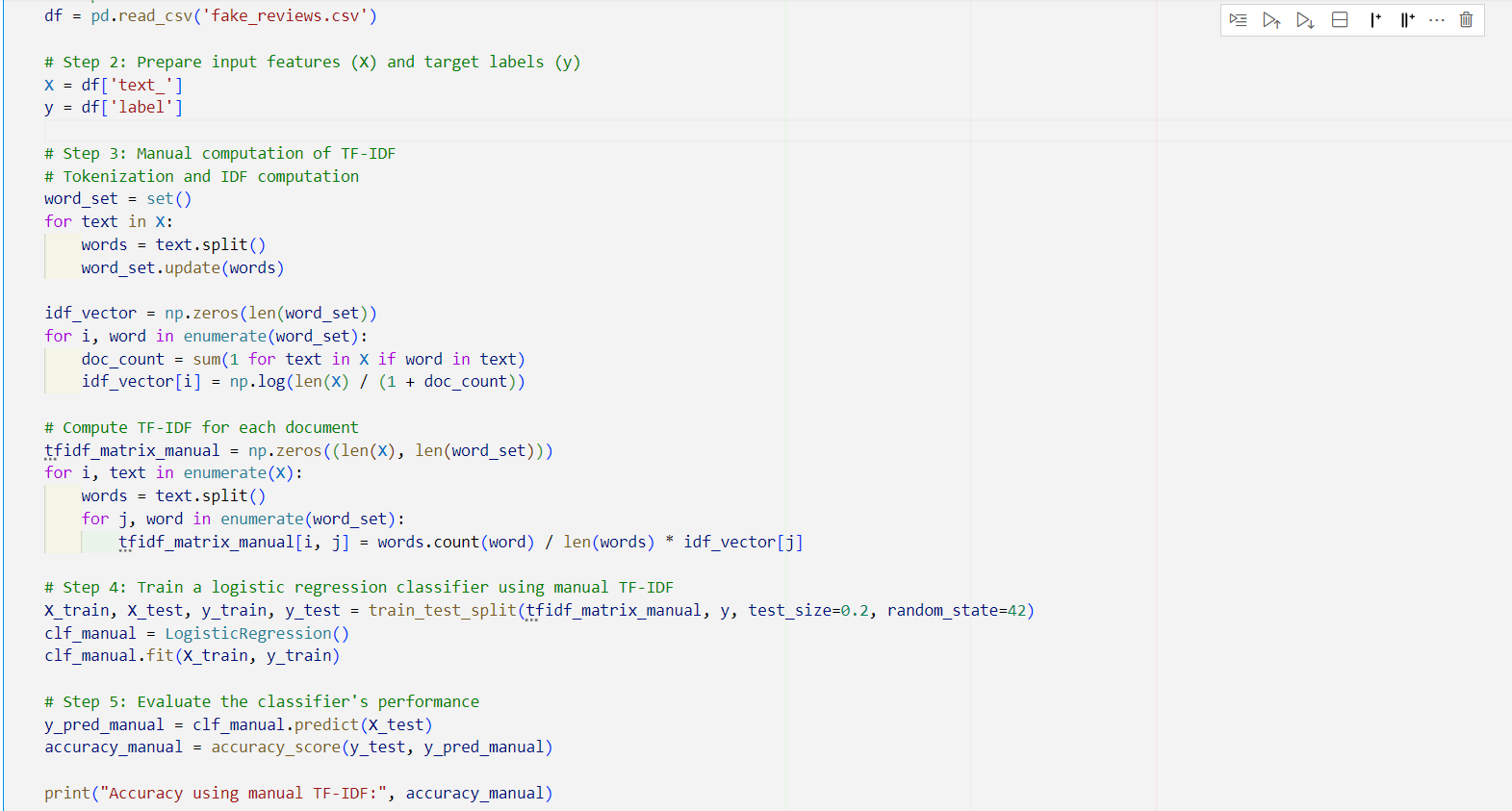


**5. Implementation of wordnet**



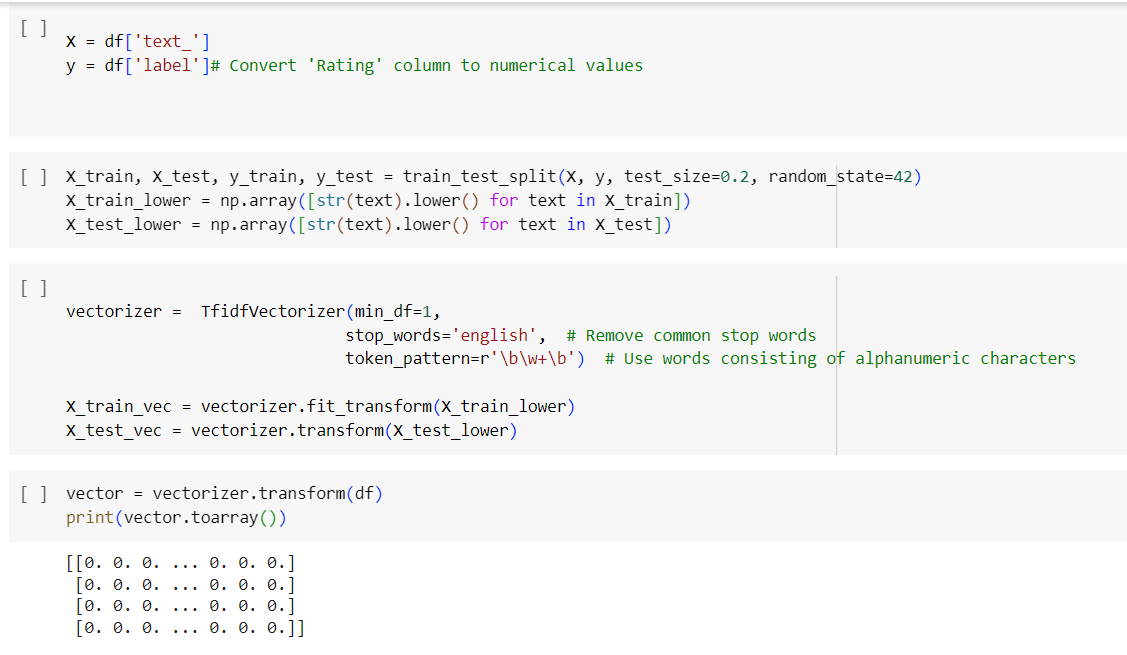
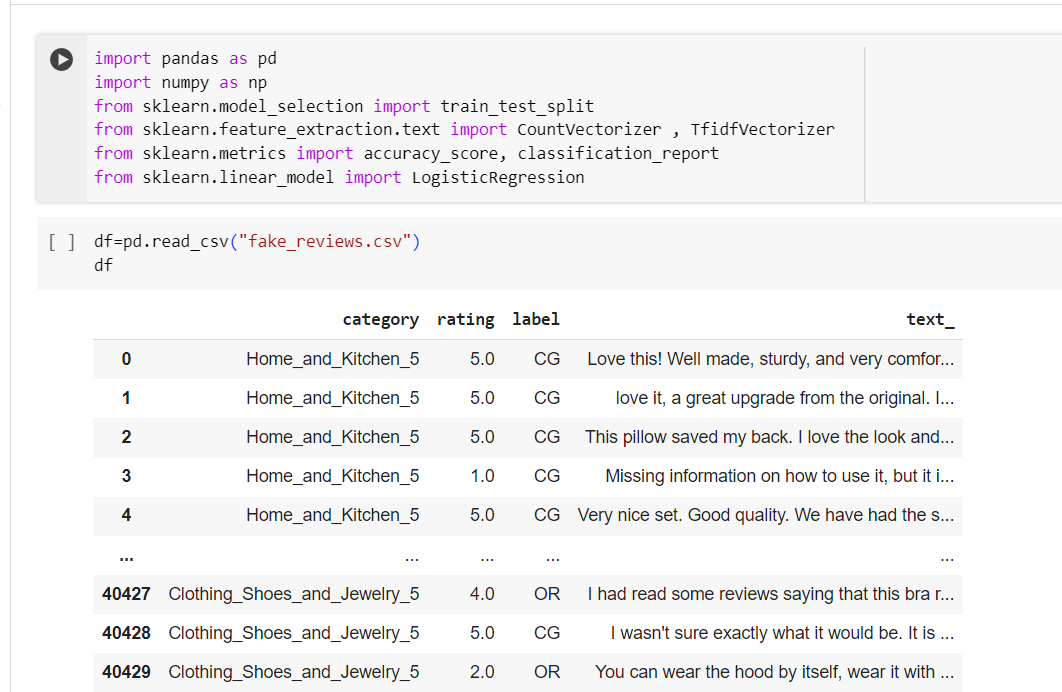
**6. Implementation of TF/TDF using manual computation and using TFldfvectorizer**

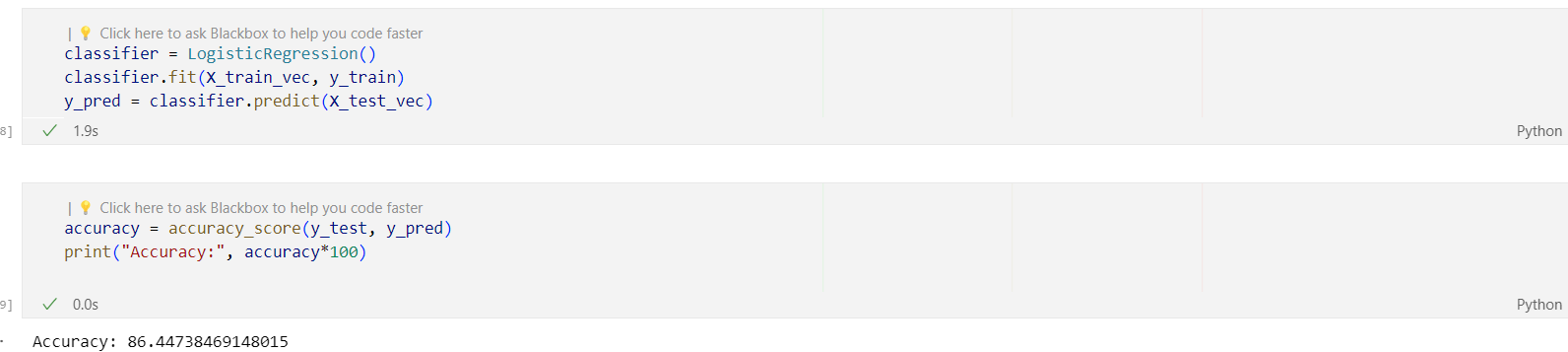
****

****

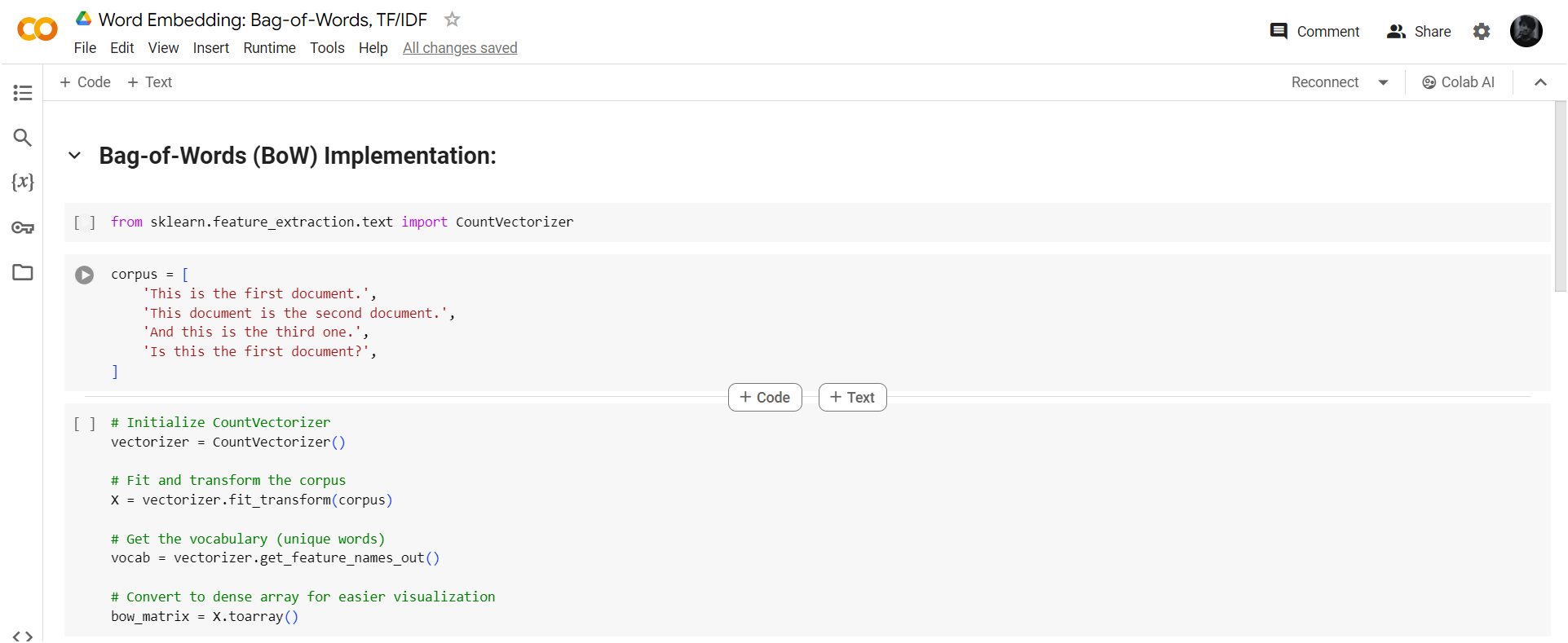
****

**Tf-IDFVectorizer**



****

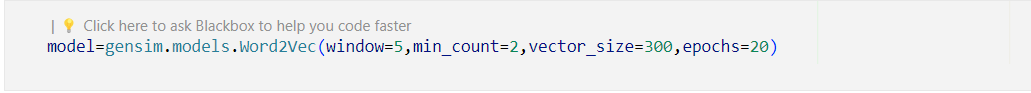
**7. Word Embedding: Bag-of-Words, TF/IDF**



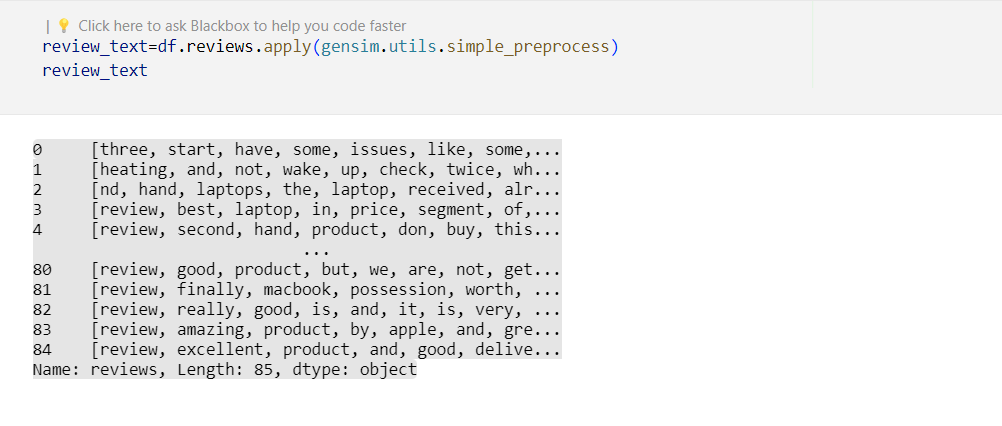
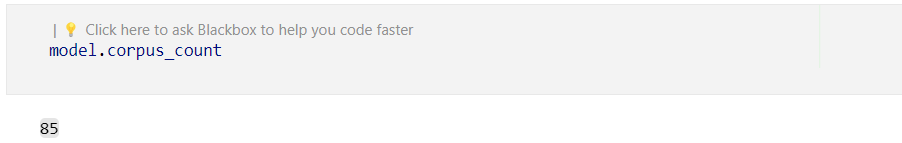
8. Word Embedding: Glove, BERT, Word2Vec

Word2Vec







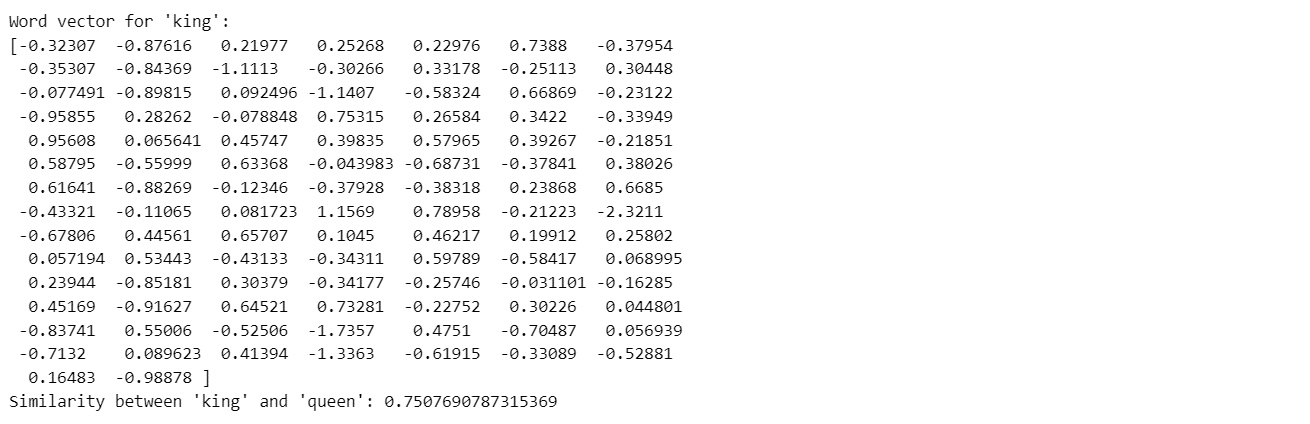
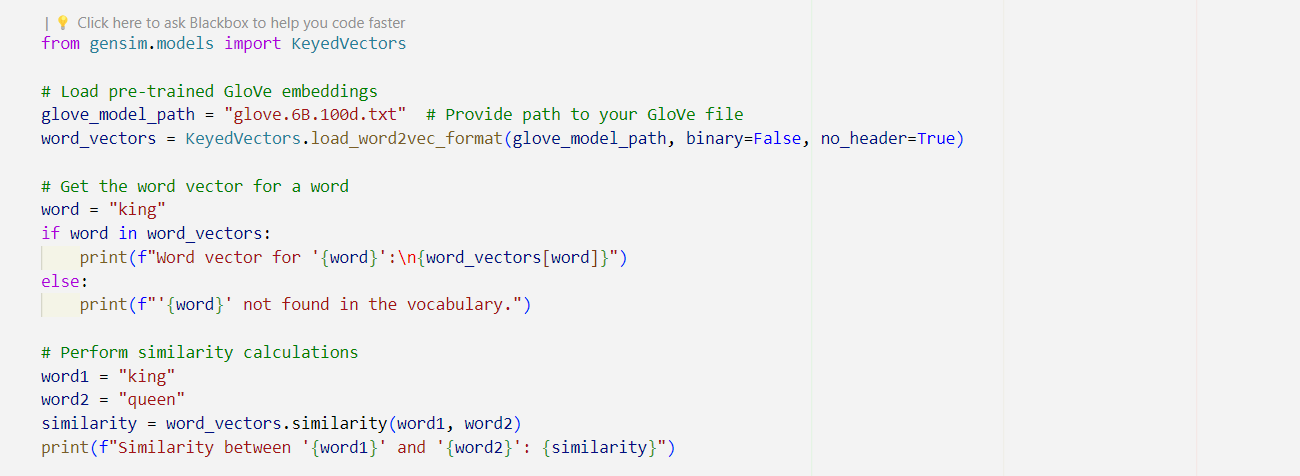




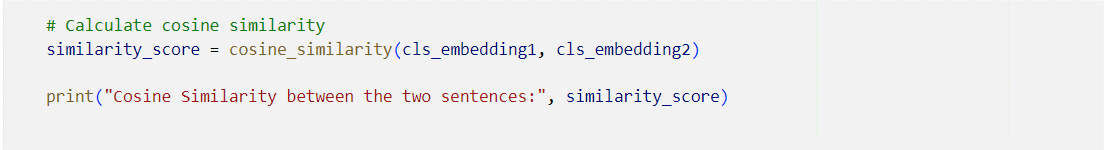
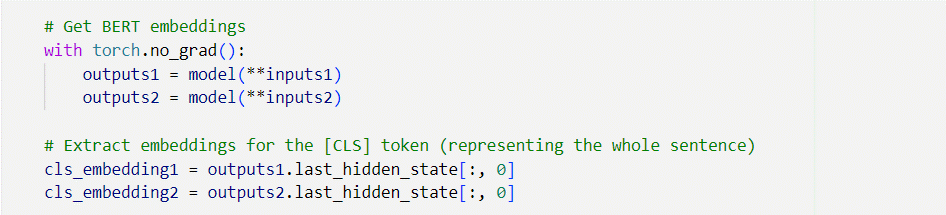
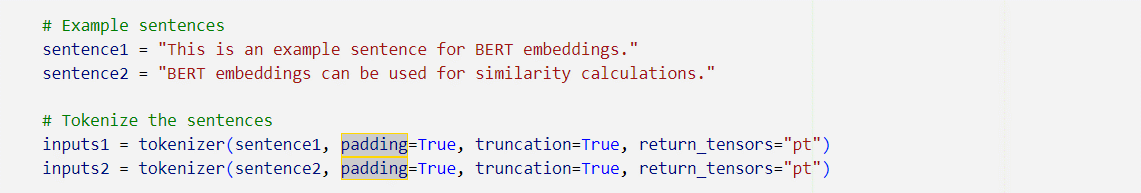
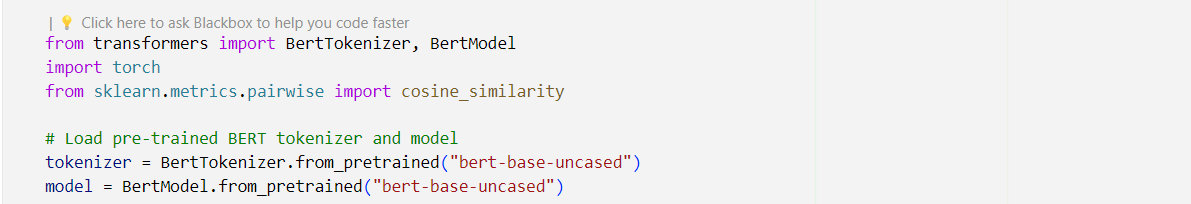




Glove

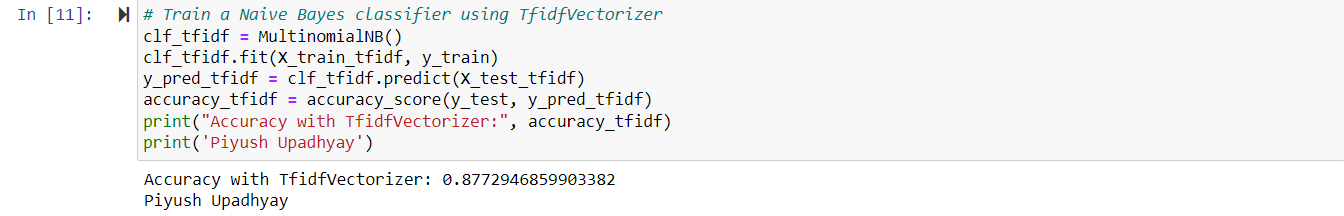


BERT

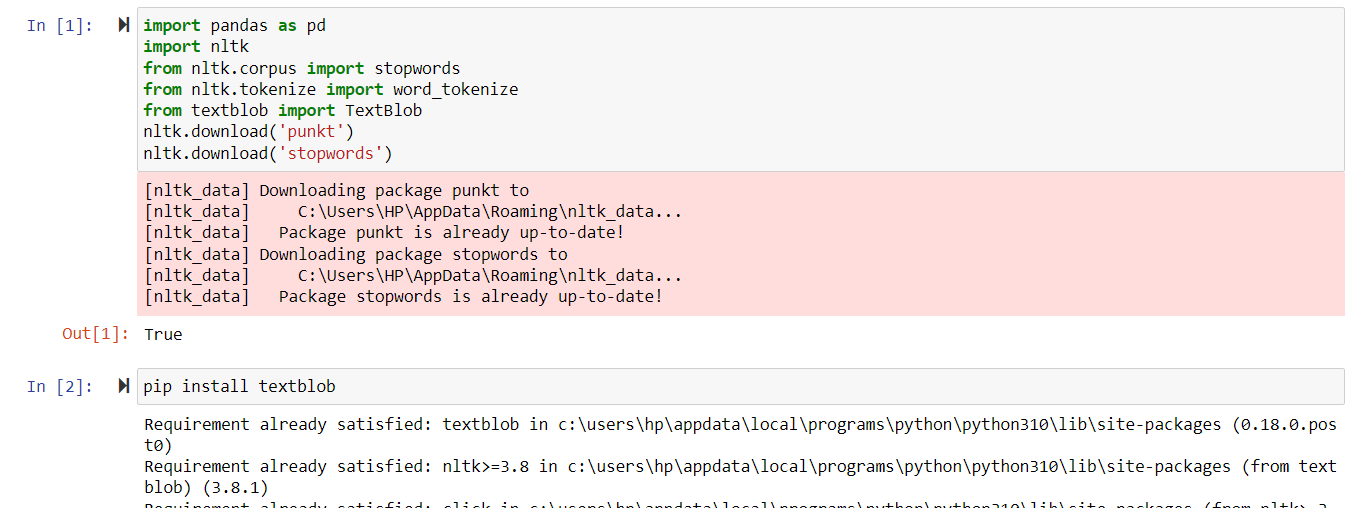


**9.Implementation of Spam/Not Spam email using TFIdfvectorizer and CountVectorizer**

****

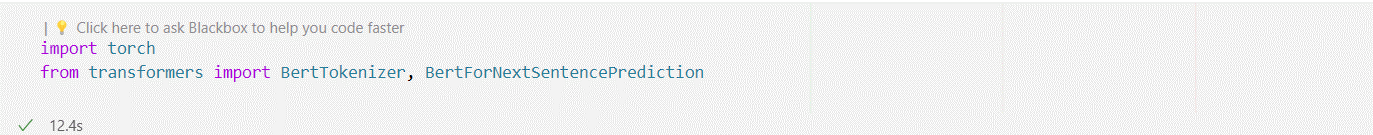


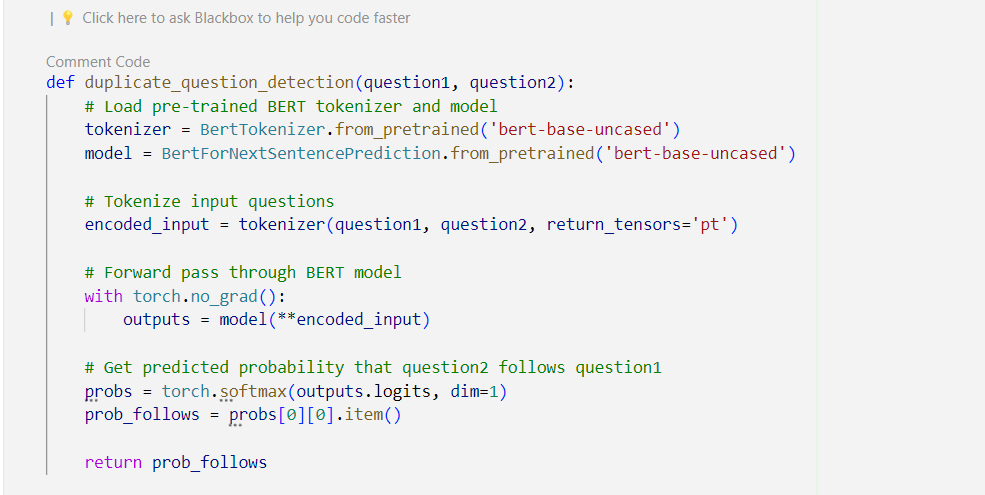
**10. Sentiment Analysis**

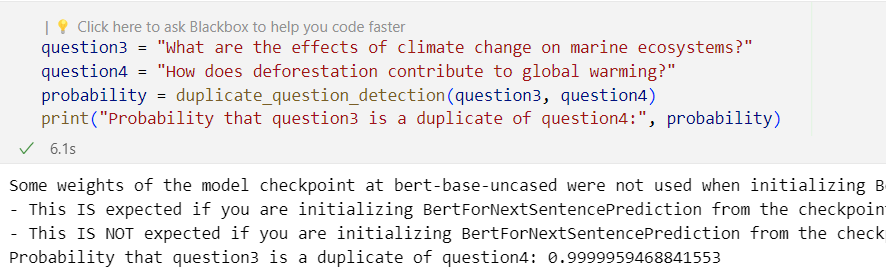
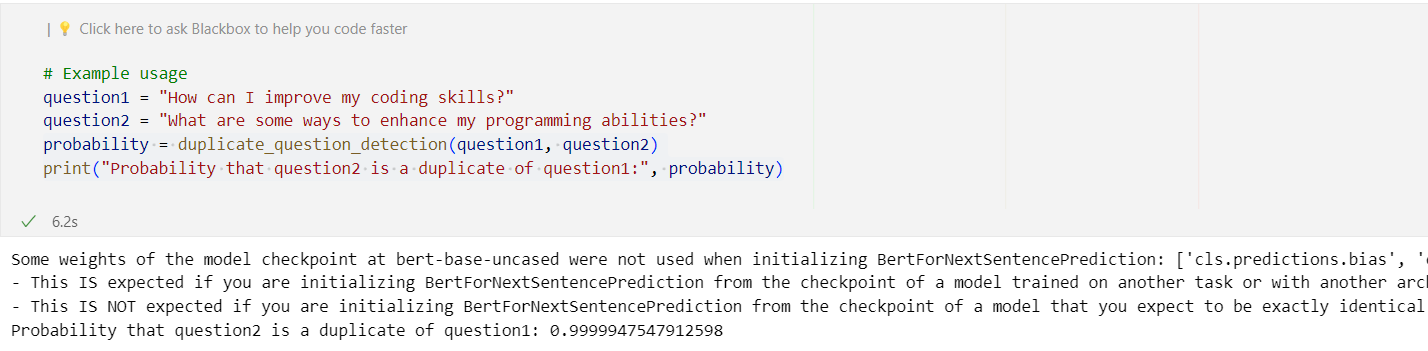




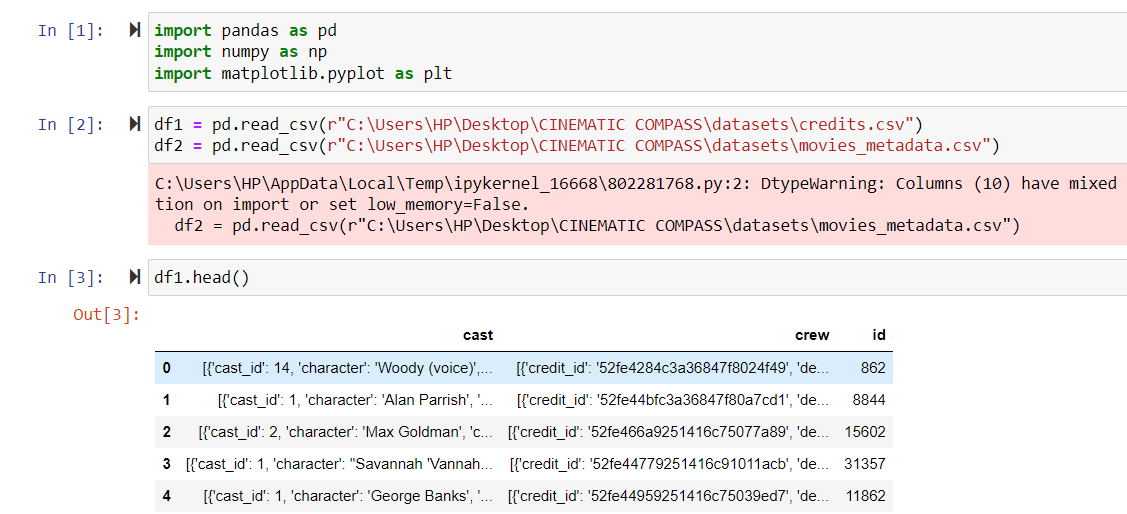
1. **Duplicate Question Detection using BERT**







1. **Recommendation System**

****





