**BOOK GENRE CLASSIFICATION BY USING BAG OF WORDS**

**AJITH MADALA**

**MANIDEEP REDDY**

**SATISH BABU NALAJALA**

**SAI PHANINDRA CHAMALLAMUDI**

**Introduction:**

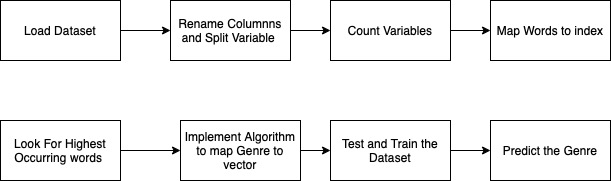
Now a days millions of books were published everywhere it hard to classify the genre of the book by reading the books manually So we designed this project using the bag of words which we taught in the class and we implemented in this project.

This way we can reduce the time of classifying the book manually and we can pick the book fastly as we wish and what we want to read.

**Model:**

Book classification based on Genre for Easy Recognition of the book using Bag of words Algorithm.

**Architecture Diagram:**

****

Load the dataset and we will remove the special symbols present in the sentences and rename the columns and we will split the variables accordingly.

Count the variables which are repeated in the sentences.

Sort the words and the Map the words to the index.

Write the function that Convert the title to vectors.

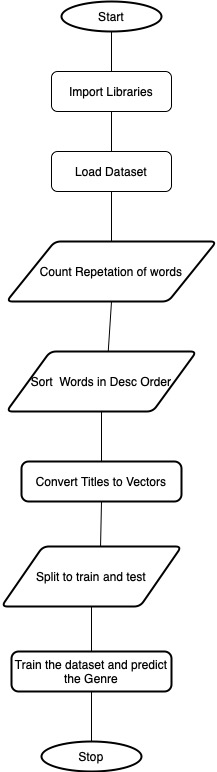
Implement an algorithm For mapping genre to vectors

Look for the highest occurance of words.

Split the data into testing and training

Predict the genre

**Work Flow Diagram:**

****

First we will import the libraries which we want for the project

We will upload the dataset which is a book.

We will count the repetition of the words in the dataset

Rename the columns and split the variable

Remove the symbols in the dataset

Count the variables and sort the words in the descending order

Covert the titles to vectors and map the words to the index  
  
Look for the highest occuring the words

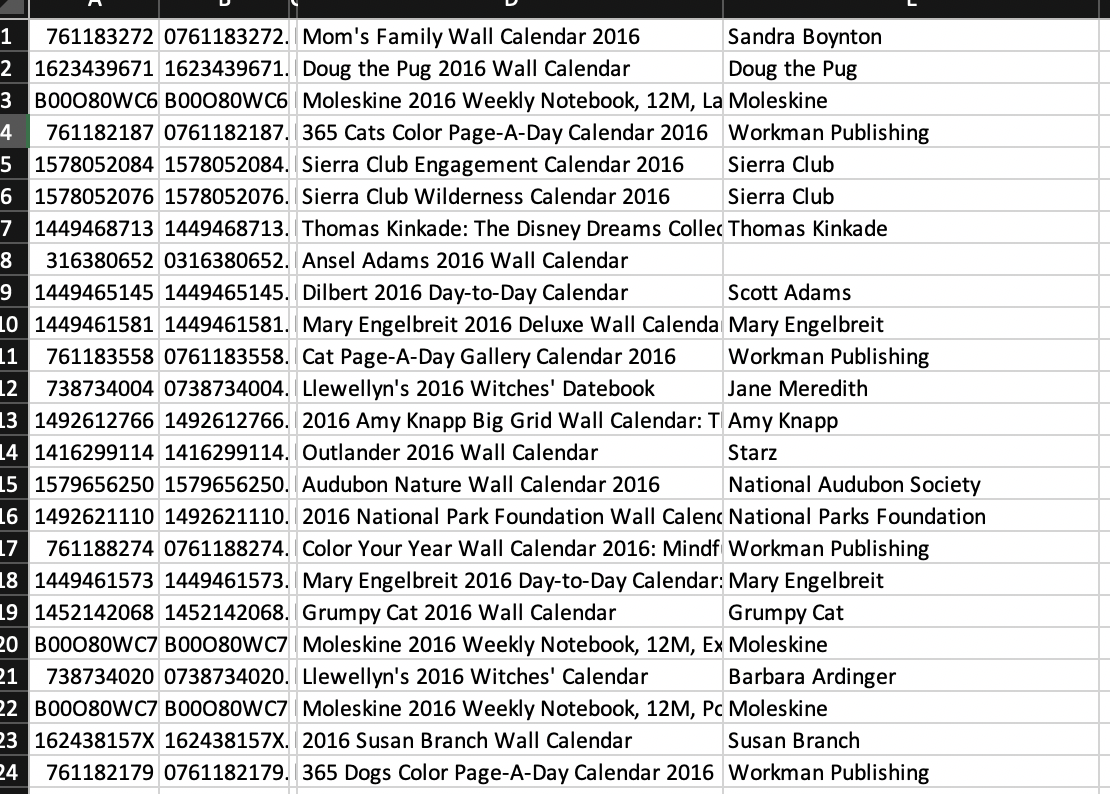
Implement the algorithm to map the genre to the vector

**Description of Dataset:**

We will take more than 10000 books in online so we can save the space in our local pc and this contains the links of the books, author, Title and categories.

From that we will take the words and divide the words into chunks and we will apply the bag of words algorithm for classifying the book.

**Design of features with diagram**

****

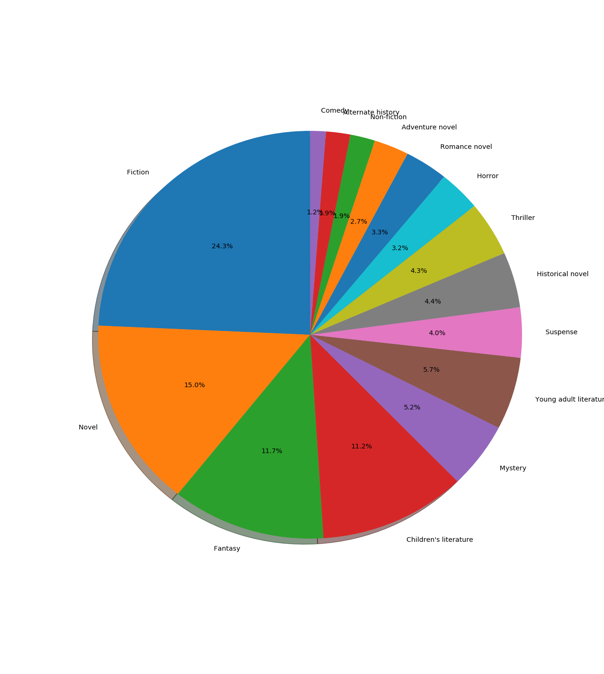
Here we will consider the title of the book, sub categories, author, SlNo, and We assigned a unique number for each and every book for easy recognisation .

**Analysis of Data**

**Data Preprocessing:**

We will Rename the columns and split the variables and remove the punctuations if available and we have to count the number of variables and map the words to index , then look for highest occurring words in the whole sentence and classify the data in the descending order.

**Graph Model**

****

By completing the classification of the book we can get the genre of the books and we will represent the result of genre in the pie chart.

**Algorithm(Bag of Words):**

Bag of Words is a way of showing the data of text when modeling text with machine learning algorithms.

If we have text data we cannot use for the model directly , we have to convert the data into vectors, we have to exclude the punctuations if needed.

During the text pre processing we have to convert the sentences to lowercase letters because if the same words get repeated that word will be considered as a different word and the frequency of the words get disturbed.

We can apply Lemmatization and stemming for the words if we need.

We have to apply stop words like ‘are , is , and, a ’,because these words are not important for the sentence these are just to connect the sentence it will not play an important role in the sentence.

After applying the stop words we will get the keywords present in the sentence.

If we take a word we will count the number of times the words got repeated and we will increase the frequency and we have to sort all the result in descending order and then we will give the importance to the words.

In Binary Bag of words the whole keywords present in the sentence and frequency were considered in the form of a table

**Disadvantage of Bag of Words:**

We cannot get the main keyword like ‘good’ the bag of words will consider the words according to the preference and it will give equal weightage for all the words present in the sentence, so that we cant do the sentimental analysis for the sentence.

**Implementation:**

Convert the Text to vectors and remove the punctuation .We will Count the repetition of the words in dataset and sort the data into Descending order.

We have to Write a function that will convert the titles into vectors.

Write an Algorithm for mapping genre to vectors and split the data and we will test and train the data.

We have to build a model.

Then we will predict the genre of book by using the bag of words Algorithm.

Results:



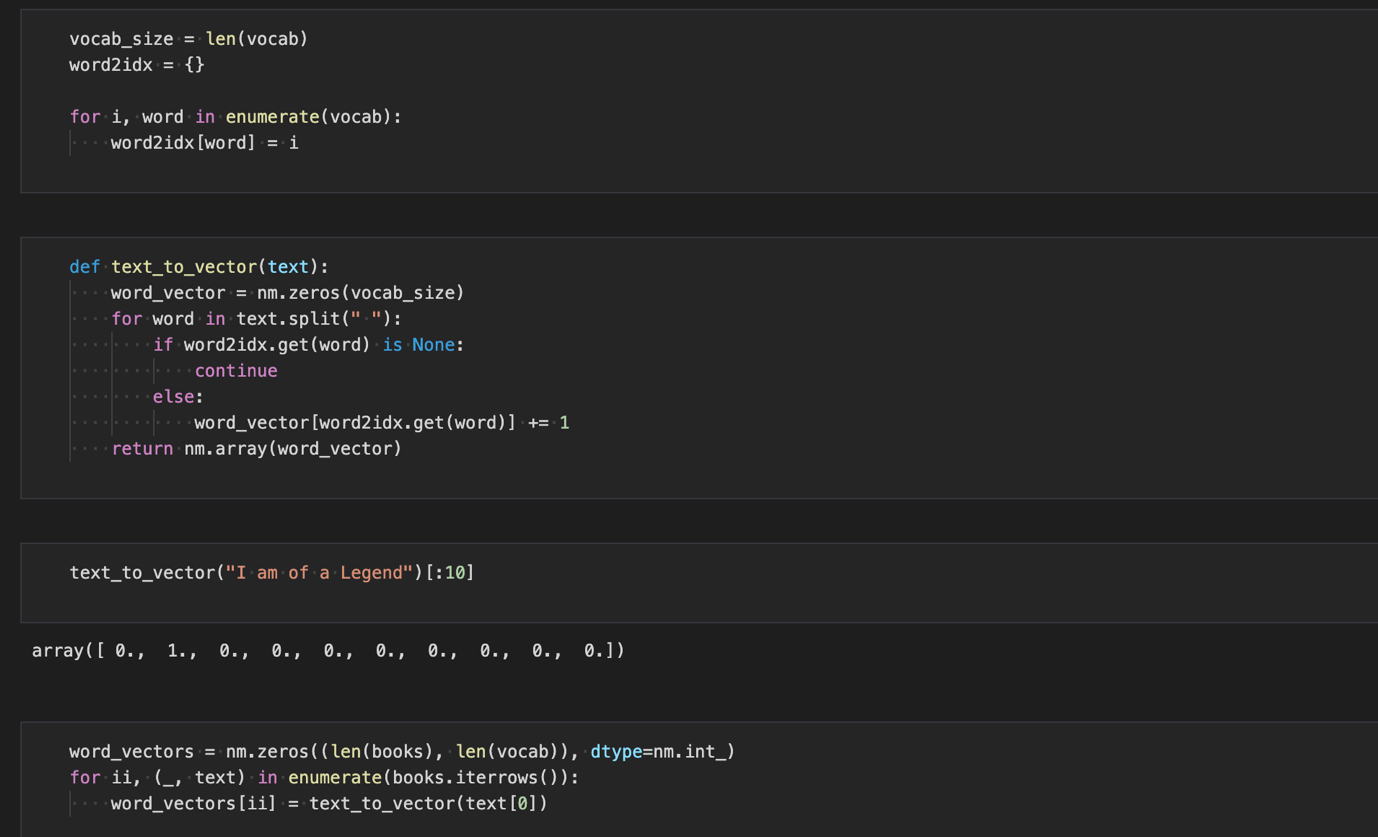
We will get the genre of the book for the given book Which will be easy to pick the book



Importing the libraries



Count the words in the dataset

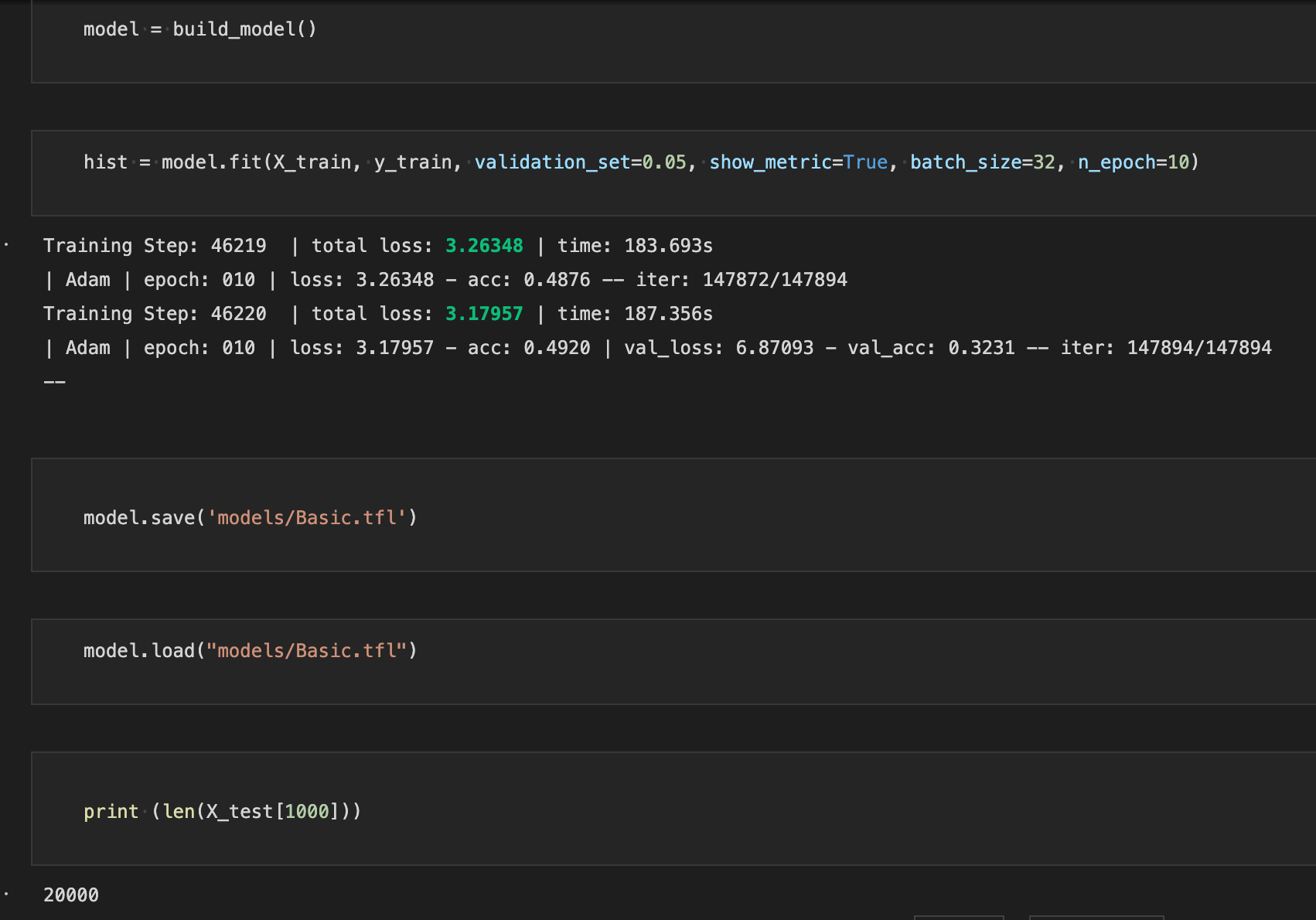
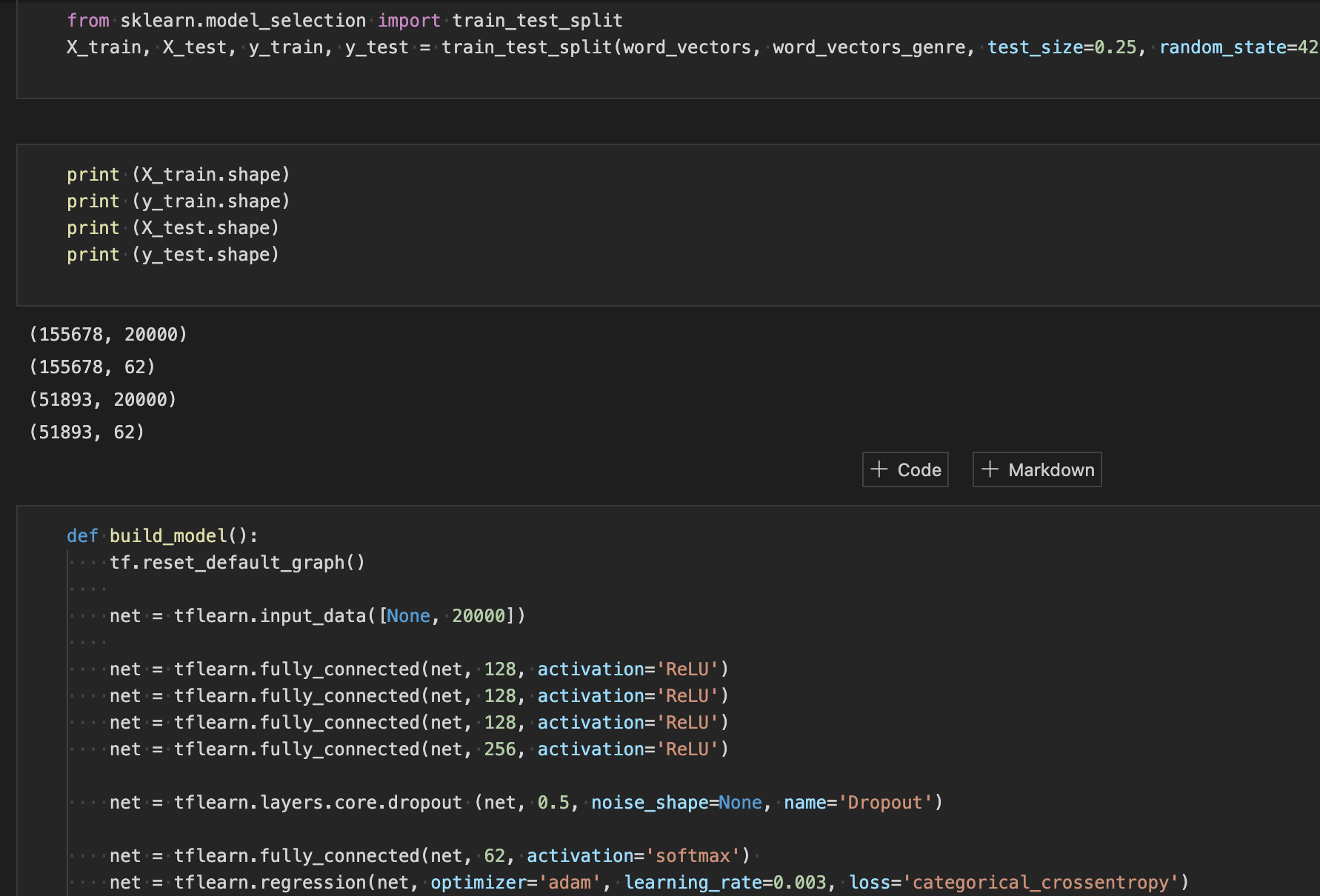
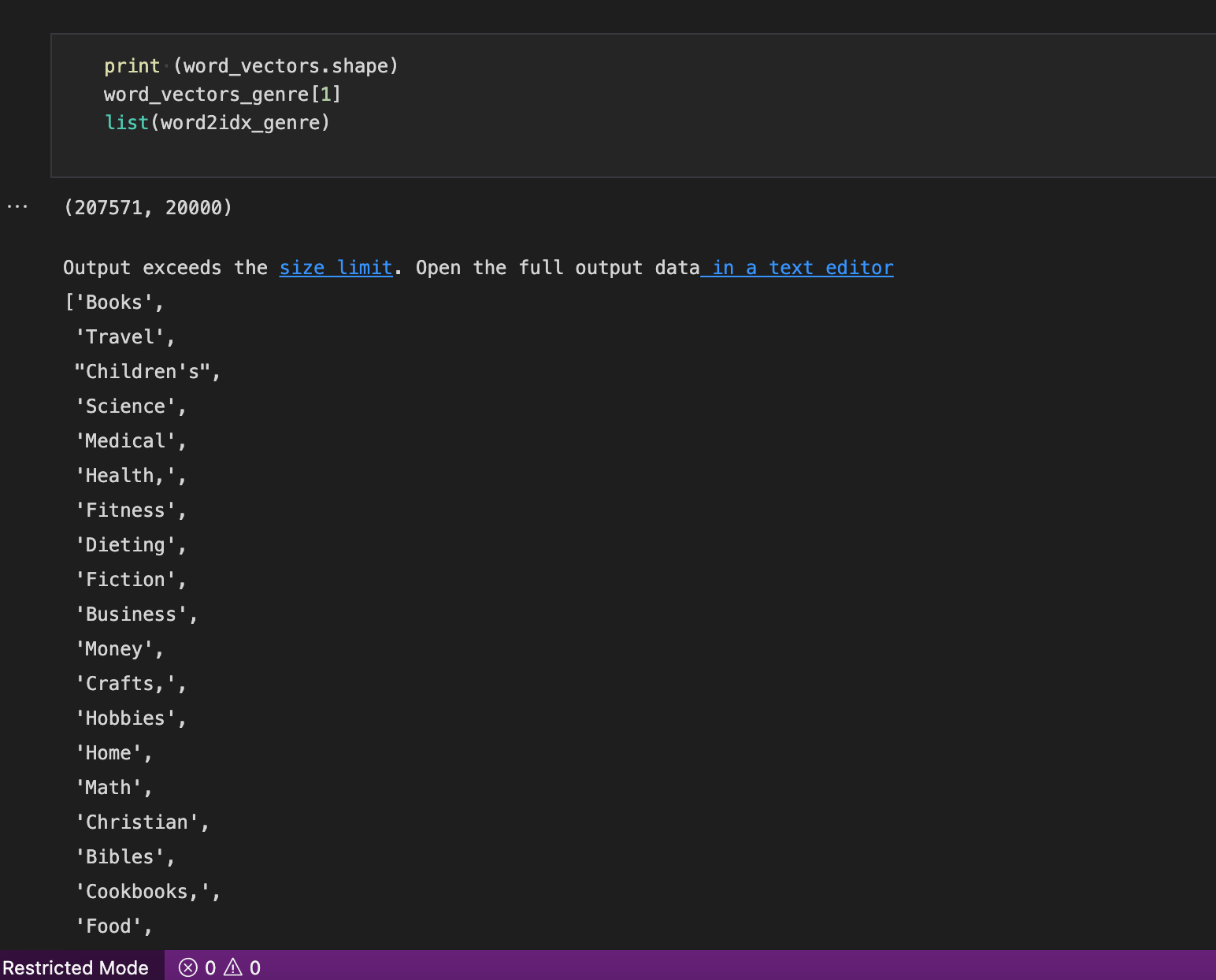


Converting Words to Vectors

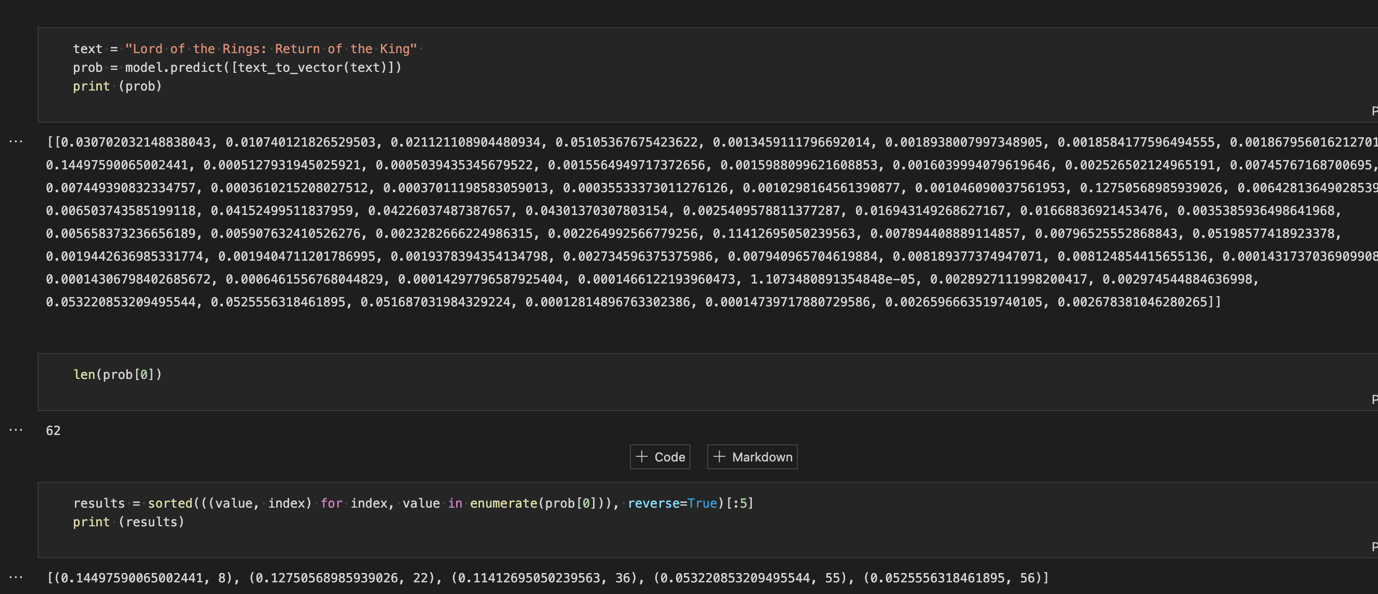


Sort the data in descending order





Training the Model





We got the desired output.

**Work Completed:**

We have obtained the desired output what we planned to Do.

**Responsibility**

AJITH MADALA. ------------- Generating valid output

MANIDEEP REDDY GADHE. ---------- processing data

SATISH BABU NALAJALA ---------- Gathering data

VENKATA SAI PHANINDRA ------understanding dataset

**Contribution:**

AJITH MADALA. ---------- 9/30

MANIDEEP REDDY GADHE---------- 9/30

SATISH BABU NALAJALA. ------------ 7/30

VENKATA SAI PHANINDRA CHAMALLAMUDI ---------- 5/30

**References**:

<https://booksummaryclub.com/genres-of-books/>

<https://www.analyticsvidhya.com/blog/2019/04/predicting-movie-genres-nlp-muli-label-classification/>

**Git Hub Link:**

https://github.com/AjithMadala/NLP-Project.git