

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt
import re
import nltk
from collections import Counter
from sklearn.feature_extraction.text import TfidfVectorizer,CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB,MultinomialNB
from sklearn.svm import SVC
from sklearn import metrics
```

```
In [2]: books = pd.read_csv('C:/Users/HP/Documents/GitHub/BookClassification/BooksDataSet.csv')
books.head()
```

Out[2]:

	Unnamed: 0	book_id	book_name	genre	summary
0	0	3248537	Drowned Wednesday	Fantasy	Drowned Wednesday is the first Trustee among ...
1	1	27796919	The Lost Hero	Fantasy	As the book opens, Jason awakens on a school ...
2	2	3910776	The Eyes of the Overworld	Fantasy	Cugel is easily persuaded by the merchant Fia...
3	3	5969644	Magic's Promise	Fantasy	The book opens with Herald-Mage Vanyel return...
4	4	3173445	Taran Wanderer	Fantasy	Taran and Gurgi have returned to Caer Dallben...

```
In [3]: # removing the unnamed : 0

books = books[['book_id', 'book_name', 'genre', 'summary']]
books.head(3)
```

Out[3]:

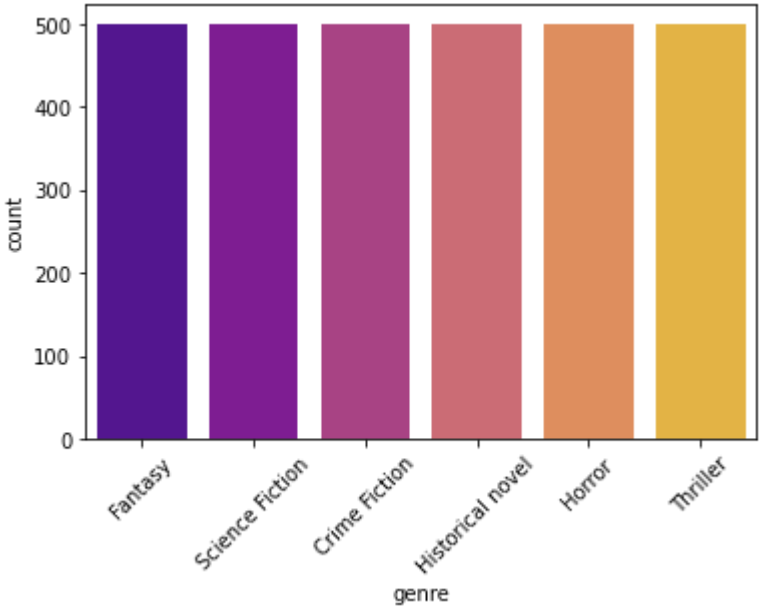
	book_id	book_name	genre	summary
0	3248537	Drowned Wednesday	Fantasy	Drowned Wednesday is the first Trustee among ...
1	27796919	The Lost Hero	Fantasy	As the book opens, Jason awakens on a school ...
2	3910776	The Eyes of the Overworld	Fantasy	Cugel is easily persuaded by the merchant Fia...

```
In [4]: books.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3000 entries, 0 to 2999
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   book_id     3000 non-null   int64
1   book_name   3000 non-null   object
2   genre       3000 non-null   object
3   summary     3000 non-null   object
dtypes: int64(1), object(3)
memory usage: 93.9+ KB
```

In [5]: `sn.countplot(x = books['genre'],palette='plasma')
plt.xticks(rotation = '45')`

Out[5]: `(array([0, 1, 2, 3, 4, 5]),
[Text(0, 0, 'Fantasy'),
Text(1, 0, 'Science Fiction'),
Text(2, 0, 'Crime Fiction'),
Text(3, 0, 'Historical novel'),
Text(4, 0, 'Horror'),
Text(5, 0, 'Thriller')])`



In [6]: `books['summary'].iloc[1]`

Out[6]: ' As the book opens, Jason awakens on a school bus, unable to remember who or where he is, or anything about his past. He is sitting next to Piper McLean and Leo Valdez, who call him by name and say they are his girlfriend and best friend respectively. All three are part of a class field trip to the Grand Canyon, and after they arrive, a classmate Dylan turns into a Venti (Storm Spirit) and attacks the trio and their trip leader, Coach Gleeson Hedge. In the ensuing fight, Jason surprises everyone, including himself, when one of his coins turns into a sword which he uses to battle the storm spirits. Coach Hedge, who reveals himself to be a satyr during the fight, is taken captive by a fleeing spirit. After the battle, a flying chariot arrives to rescue the trio, but one of the people in it, Annabeth, is upset when she discovers that her missing boyfriend, Percy Jackson, is not there as she expected. Annabeth, seeking Percy, was told in a vision from the goddess Hera to look there for the "guy with one shoe", but this turns out to be Jason, who had a shoe destroyed during the fight. Jason, Piper, and Leo are told that they are demigods and are taken back to Camp Half-Blood where they meet other Greek demigod children like themselves. There, Leo is revealed as a son of Hephaestus, Piper as a daughter of Aphrodite and Jason as a son of Zeus, though Hera tells him he is her champion. Jason later discovers that he is the full brother of Zeus's demigod daughter Thalia Grace, who is a Hunter of Artemis. Shortly after they arrive, the three are given a quest to rescue Hera, who has been captured, and they set off. They soon discover that their enemies are working under orders from Gaea to overthrow the gods. During their quest, they encounter Thalia and the Hunters, who have been looking for Percy. Thalia and Jason reunite for the first since Jason was captured at the age of two. On the way to Aeolus's castle, Jason, Leo and Piper become separated from Thalia, who promises to meet them at the Wolf House, the last place Thalia had seen Jason before this meeting. After being nearly apprehended by Aeolus, who is under Gaea's orders, the trio manage to escape thanks to Mellie, Aeolus's former assistant, and end up in San Francisco, thanks to the result of a dream Piper had with Aphrodite. After landing in San Francisco, the trio rush to Mt. Diablo to fight the giant Enceladus, who has kidnapped Piper's father. They manage to kill the giant and save Piper's father, after which they rush to the Wolf House to free Hera. Although the heroes and the Hunters save Hera, the king of the giants, Porphyron, rises fully and disappears into a hole in the Earth. Jason's memory then starts returning, and he remembers that he is a hero from a Roman counterpart to Camp Half-Blood somewhere near San Francisco, and is the son of Jupiter, Zeus's Roman aspect. He realizes that Hera, also known as Juno, has switched him with Percy Jackson, who will be at the Roman camp with no memory of his life, in the hopes that the two camps would ultimately work together to fight the giants and defeat the goddess Gaea.'

In [7]:

```
## cleaning the text

def cleantext(text):

    # removing the "\"

    text = re.sub("'\\'", "", text)

    # removing special symbols

    text = re.sub("[^a-zA-Z]", " ", text)

    # removing the whitespaces

    text = ' '.join(text.split())

    # convert text to lowercase

    text = text.lower()

    return text

books['summary'] = books['summary'].apply(lambda x:cleantext(x))
books['summary'].iloc[1]
```

Out[7]:

'as the book opens jason awakens on a school bus unable to remember who or where he is or anything about his past he is sitting next to piper mclean and leo valdez who call him by name and say they are his girlfr iend and best friend respectively all three are part of a class field trip to the grand canyon and after they arrive a classmate dylan turns into a venti storm spirit and attacks the trio and their trip leader co ach gleeson hedge in the ensuing fight jason surprises everyone including himself when one of his coins turns into a sword which he uses to battle the storm spirits coach hedge who reveals himself to be a satyr d uring the fight is taken captive by a fleeing spirit after the battle a flying chariot arrives to rescue the trio but one of the people in it annabeth is upset when she discovers that her missing boyfriend percy jackson is not there as she expected annabeth seeking percy was told in a vision from the goddess hera to look there for the guy with one shoe but this turns out to be jason who had a shoe destroyed during the fi ght jason piper and leo are told that they are demigods and are taken back to camp half blood where they meet other greek demigod children like themselves there leo is revealed as a son of hephaestus piper as a d aughter of aphrodite and jason as a son of zeus though hera tells him he is her champion jason later discovers that he is the full brother of zeus s demigod daughter thalia grace who is a hunter of artemis shortl y after they arrive the three are given a quest to rescue hera who has been captured and they set off they soon discover that their enemies are working under orders from gaea to overthrow the gods during their qu est they encounter thalia and the hunters who have been looking for percy thalia and jason reunite for the first since jason was captured at the age of two on the way to aeolus s castle jason leo and piper become separated from thalia who promises to meet them at the wolf house the last place thalia had seen jason before this meeting after being nearly apprehended by aeolus who is under gaea s orders the trio manage to es cape thanks to mellie aeolus s former assistant and end up in san francisco thanks to the result of a dream piper had with aphrodite after landing in san francisco the trio rush to mt diablo to fight the giant en celadus who has kidnapped piper s father they manage to kill the giant and save piper s father after which they rush to the wolf house to free hera although the heroes and the hunters save hera the king of the gi ants porphyron rises fully and disappears into a hole in the earth jason s memory then starts returning and he remembers that he is a hero from a roman counterpart to camp half blood somewhere near san francisco and is the son of jupiter zeus s roman aspect he realizes that hera also known as juno has switched him with percy jackson who will be at the roman camp with no memory of his life in the hopes that the two camps would ultimately work together to fight the giants and defeat the goddess gaea'

Plotting the most Frequent Words

In [8]:

```
def showmostfrequentwords(text,no_of_words):

    allwords = ' '.join([char for char in text])
    allwords = allwords.split()
    fdist = nltk.FreqDist(allwords)

    wordsdf = pd.DataFrame({'word':list(fdist.keys()),'count':list(fdist.values())})

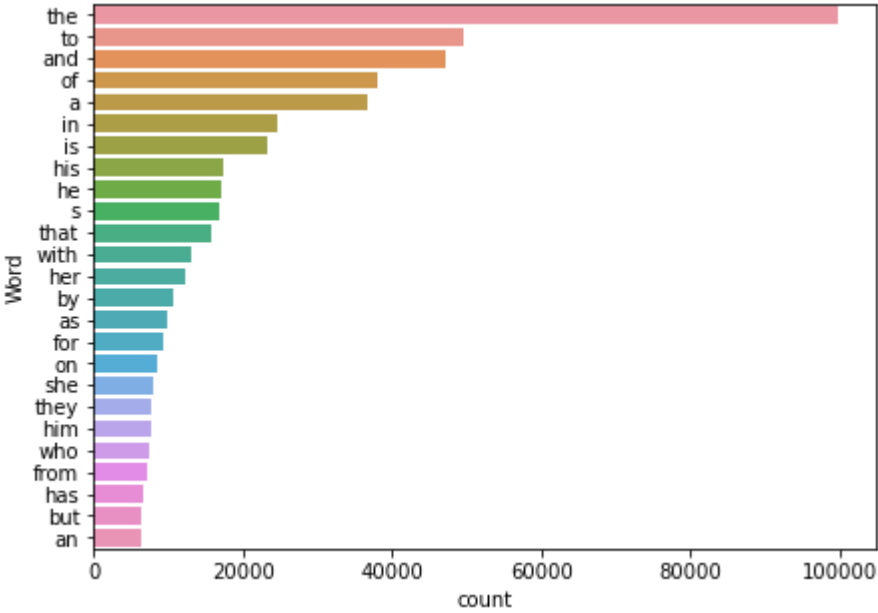
    df = wordsdf.nlargest(columns="count",n = no_of_words)

    plt.figure(figsize=(7,5))
    ax = sn.barplot(data=df,x = 'count',y = 'word')
    ax.set(ylabel = 'Word')
    plt.show()

    return wordsdf

# 25 most frequent words

wordsdf = showmostfrequentwords(books['summary'],25)
```



In [9]:

```
wordsdf.sort_values('count',ascending=False).head(10).style.background_gradient(cmap = 'plasma')
```

Out[9]:

	word	count
3	the	99808
17	to	49665
14	and	47183
46	of	38003
23	a	36749
80	in	24496
2	is	23358
49	his	17464
51	he	17210

word	count

Removing the stopwords

In [10]:

```

nltk.download('stopwords')
from nltk.corpus import stopwords
stop_words = set(stopwords.words('english'))

# removing the stopwords

def removestopwords(text):

    removedstopword = [word for word in text.split() if word not in stop_words]
    return ' '.join(removedstopword)

books['summary'] = books['summary'].apply(lambda x:removestopwords(x))
books['summary'].iloc[1]
```

[nltk_data] Error loading stopwords: <urlopen error [Errno 11001]
[nltk_data] getaddrinfo failed>

Out[10]:

'book opens jason awakens school bus unable remember anything past sitting next piper mclean leo valdez call name say girlfriend best friend respectively three part class field trip grand canyon arrive classmate dylan turns venti storm spirit attacks trio trip leader coach gleeson hedge ensuing fight jason surprises everyone including one coins turns sword uses battle storm spirits coach hedge reveals satyr fight taken captive fleeing spirit battle flying chariot arrives rescue trio one people annabeth upset discovers missing boyfriend percy jackson expected annabeth seeking percy told vision goddess hera look guy one shoe turns jason shoe destroyed fight jason piper leo told demigods taken back camp half blood meet greek demigod children like leo revealed son hephaestus piper daughter aphrodite jason son zeus though hera tells champion jason later discovers full brother zeus demigod daughter thalia grace hunter artemis shortly arrive three given quest rescue hera captured set soon discover enemies working orders gaea overthrow gods quest encounter thalia hunters looking percy thalia jason reunite first since jason captured age two way aeolus castle jason leo piper become separated thalia promises meet wolf house last place thalia seen jason meeting nearly apprehended aeolus gaea orders trio manage escape thanks mellie aeolus former assistant end san francisco thanks result dream piper aphrodite landing san francisco trio rush mt diablo fight giant enceladus kidnapped piper father manage kill giant save piper father rush wolf house free hera although heroes hunters save hera king giants porphyron rises fully disappears hole earth jason memory starts returning remember s hero roman counterpart camp half blood somewhere near san francisco son jupiter zeus roman aspect realizes hera also known juno switched percy jackson roman camp memory life hopes two camps would ultimately work together fight giants defeat goddess gaea'

Lemmatizing

In [17]:

```

import nltk
##nltk.download('wordnet')
nltk.download('all')
from nltk.stem import WordNetLemmatizer

lemma=WordNetLemmatizer()

def lematizing(sentence):
    stemSentence = ""
    for word in sentence.split():
        stem = lemma.lemmatize(word)
        stemSentence += stem
        stemSentence += " "
    stemSentence = stemSentence.strip()
    return stemSentence

books['summary'] = books['summary'].apply(lambda x: lematizing(x))
```

[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package abc is already up-to-date!
[nltk_data] | Downloading package alpino to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package alpino is already up-to-date!

```
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package averaged_perceptron_tagger is already up-
[nltk_data] | to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_ru to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_ru is already
[nltk_data] | up-to-date!
[nltk_data] | Downloading package basque_grammars to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package basque_grammars is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package comtrans to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
```

```
[nltk_data] | Package europarl_raw is already up-to-date!  
[nltk_data] | Downloading package extended_omw to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Unzipping corpora\extended_omw.zip.  
[nltk_data] | Downloading package floresta to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package floresta is already up-to-date!  
[nltk_data] | Downloading package framenet_v15 to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package framenet_v15 is already up-to-date!  
[nltk_data] | Downloading package framenet_v17 to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package framenet_v17 is already up-to-date!  
[nltk_data] | Downloading package gazetteers to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package gazetteers is already up-to-date!  
[nltk_data] | Downloading package genesis to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package genesis is already up-to-date!  
[nltk_data] | Downloading package gutenber to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package gutenber is already up-to-date!  
[nltk_data] | Downloading package ieer to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package ieer is already up-to-date!  
[nltk_data] | Downloading package inaugural to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Unzipping corpora\inaugural.zip.  
[nltk_data] | Downloading package indian to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package indian is already up-to-date!  
[nltk_data] | Downloading package jeita to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package jeita is already up-to-date!  
[nltk_data] | Downloading package kimmo to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package kimmo is already up-to-date!  
[nltk_data] | Downloading package knbc to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package knbc is already up-to-date!  
[nltk_data] | Downloading package large_grammars to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package large_grammars is already up-to-date!  
[nltk_data] | Downloading package lin_thesaurus to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package lin_thesaurus is already up-to-date!  
[nltk_data] | Downloading package mac_morpho to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package mac_morpho is already up-to-date!  
[nltk_data] | Downloading package machado to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package machado is already up-to-date!  
[nltk_data] | Downloading package masc_tagged to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package masc_tagged is already up-to-date!  
[nltk_data] | Downloading package maxent_ne_chunker to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package maxent_ne_chunker is already up-to-date!  
[nltk_data] | Downloading package maxent_treebank_pos_tagger to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package maxent_treebank_pos_tagger is already up-  
[nltk_data] | to-date!  
[nltk_data] | Downloading package moes_sample to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package moes_sample is already up-to-date!  
[nltk_data] | Downloading package movie_reviews to  
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Package movie_reviews is already up-to-date!
```



```
[nltk_data] | Downloading package mte_teip5 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package mte_teip5 is already up-to-date!
[nltk_data] | Downloading package mwa_ppdb to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package mwa_ppdb is already up-to-date!
[nltk_data] | Downloading package names to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package nombank.1.0 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package nombank.1.0 is already up-to-date!
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package nonbreaking_prefixes is already up-to-date!
[nltk_data] | Downloading package nps_chat to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package nps_chat is already up-to-date!
[nltk_data] | Downloading package omw to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\omw-1.4.zip.
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package opinion_lexicon is already up-to-date!
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package panlex_swadesh is already up-to-date!
[nltk_data] | Downloading package paradigms to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package paradigms is already up-to-date!
[nltk_data] | Downloading package pe08 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\pe08.zip.
[nltk_data] | Downloading package perluniprops to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package perluniprops is already up-to-date!
[nltk_data] | Downloading package pil to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package pil is already up-to-date!
[nltk_data] | Downloading package pl196x to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package pl196x is already up-to-date!
[nltk_data] | Downloading package porter_test to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package porter_test is already up-to-date!
[nltk_data] | Downloading package ppattach to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package ppattach is already up-to-date!
[nltk_data] | Downloading package problem_reports to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package problem_reports is already up-to-date!
[nltk_data] | Downloading package product_reviews_1 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package product_reviews_1 is already up-to-date!
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package product_reviews_2 is already up-to-date!
[nltk_data] | Downloading package propbank to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package propbank is already up-to-date!
[nltk_data] | Downloading package pros_cons to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package pros_cons is already up-to-date!
[nltk_data] | Downloading package ptb to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
```



```
[nltk_data] | Package ptb is already up-to-date!
[nltk_data] | Downloading package punkt to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package qc to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package qc is already up-to-date!
[nltk_data] | Downloading package reuters to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package reuters is already up-to-date!
[nltk_data] | Downloading package rslp to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package rslp is already up-to-date!
[nltk_data] | Downloading package rte to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package rte is already up-to-date!
[nltk_data] | Downloading package sample_grammars to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package sample_grammars is already up-to-date!
[nltk_data] | Downloading package semcor to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package semcor is already up-to-date!
[nltk_data] | Downloading package senseval to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package senseval is already up-to-date!
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package sentence_polarity is already up-to-date!
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package sentiwordnet is already up-to-date!
[nltk_data] | Downloading package shakespeare to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\sinica_treebank.zip.
[nltk_data] | Downloading package smultron to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package smultron is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package spanish_grammars is already up-to-date!
[nltk_data] | Downloading package state_union to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package state_union is already up-to-date!
[nltk_data] | Downloading package stopwords to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package subjectivity to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package subjectivity is already up-to-date!
[nltk_data] | Downloading package swadesh to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package swadesh is already up-to-date!
[nltk_data] | Downloading package switchboard to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package switchboard is already up-to-date!
[nltk_data] | Downloading package tagsets to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package tagsets is already up-to-date!
[nltk_data] | Downloading package timit to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package timit is already up-to-date!
[nltk_data] | Downloading package toolbox to
```

```
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package toolbox is already up-to-date!
[nltk_data] | Downloading package treebank to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package udhr to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package udhr is already up-to-date!
[nltk_data] | Downloading package udhr2 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package udhr2 is already up-to-date!
[nltk_data] | Downloading package unicode_samples to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package unicode_samples is already up-to-date!
[nltk_data] | Downloading package universal_tagset to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package universal_tagset is already up-to-date!
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package universal_treebanks_v20 is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package vader_lexicon is already up-to-date!
[nltk_data] | Downloading package verbnet to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package verbnet is already up-to-date!
[nltk_data] | Downloading package verbnet3 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package verbnet3 is already up-to-date!
[nltk_data] | Downloading package webtext to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package webtext is already up-to-date!
[nltk_data] | Downloading package wmt15_eval to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package wmt15_eval is already up-to-date!
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package word2vec_sample is already up-to-date!
[nltk_data] | Downloading package wordnet to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\wordnet2021.zip.
[nltk_data] | Downloading package wordnet31 to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\wordnet31.zip.
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package ycoe to
[nltk_data] | C:\Users\HP\AppData\Roaming\nltk_data...
```

Stemming

In [18]:

```
# Approach

from nltk.stem import PorterStemmer
stemmer = PorterStemmer()

def stemming(sentence):

    stemmed_sentence = ""
    for word in sentence.split():
        stem = stemmer.stem(word)
        stemmed_sentence+=stem
        stemmed_sentence+=" "

    stemmed_sentence = stemmed_sentence.strip()
    return stemmed_sentence

books['summary'] = books['summary'].apply(lambda text:stemming(text))
books['summary'].iloc[1]
```

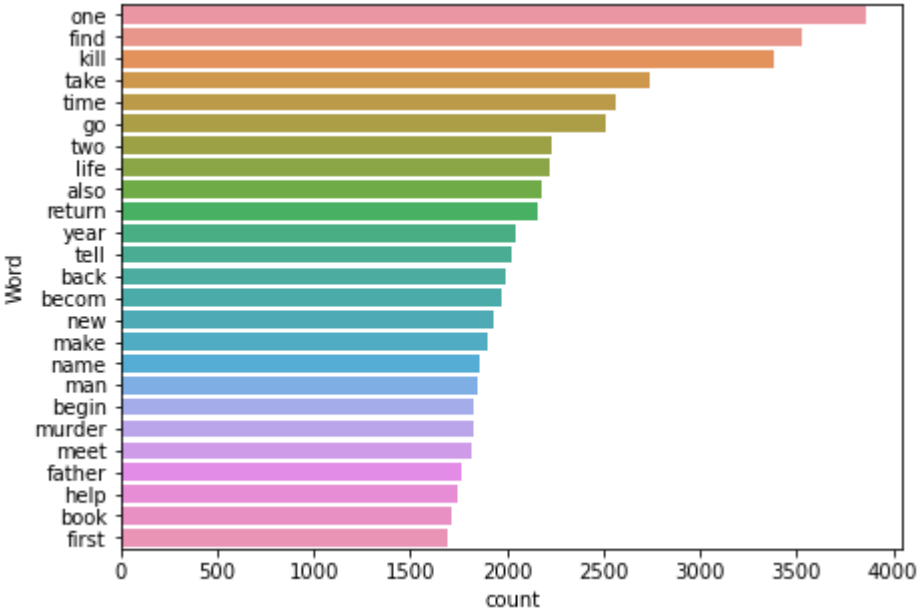
Out[18]:

'book open jason awaken school bu unabl rememb anyth past sit next piper mclean leo valdez call name say girlfriend best friend respect three part class field trip grand canyon arriv classmat dylan turn venti sto
rm spirit attack trio trip leader coach gleeson hedg ensu fight jason surpris everyon includ one coin turn sword us battl storm spirit coach hedg reveal satyr fight taken captiv flee spirit battl fli chariot arri
v rescu trio one peopl annabeth upset discov miss boyfriend perci jackson expect annabeth seek perci told vision goddess hera look guy one shoe turn jason shoe destroy fight jason piper leo told demigod taken bac
k camp half blood meet greek demigod child like leo reveal son hephaestu piper daughter aphrodit jason son zeu though hera tell champion jason later discov full brother zeu demigod daughter thalia grace hunter ar
temi shortli arriv three given quest rescu hera captur set soon discov enemi work order gaea overthrow god quest encount thalia hunter look perci thalia jason reunite first sinc jason captur age two way aeolu cast
l jason leo piper becom separ thalia promis meet wolf hous last place thalia seen jason meet nearli apprehend aeolu gaea order trio manag escap thank melli aeolu former assist end san francisco thank result dream
piper aphrodit land san francisco trio rush mt diablo fight giant enceladu kidnap piper father manag kill giant save piper father rush wolf hous free hera although hero hunter save hera king giant porphyron rise
fulli disappear hole earth jason memori start return rememb hero roman counterpart camp half blood somewher near san francisco son jupit zeu roman aspect realiz hera also known juno switch perci jackson roman cam
p memori life hope two camp would ultim work togeth fight giant defeat goddess gaea'

In [19]:

```
# visualising frewords

freq_df = showmostfrequentwords(books['summary'],25)
```



In [20]:

```
freq_df.sort_values('count',ascending=False).head(10).style.background_gradient(cmap = 'plasma')
```

Out[20]:

	word	count
242	one	3858
230	find	3533

	word	count
240	kill	3380
153	take	2743
1143	time	2563
88	go	2509
38	two	2235
418	life	2224
415	also	2179

Encoding

```
In [21]: books_list = list(books['genre'].unique())
encode = [i for i in range(len(books_list))]
mapper = dict(zip(books_list,encode))
print(mapper)
```

{'Fantasy': 0, 'Science Fiction': 1, 'Crime Fiction': 2, 'Historical novel': 3, 'Horror': 4, 'Thriller': 5}

```
In [22]: books['genre'] = books['genre'].map(mapper)
books['genre'].unique()
```

Out[22]: array([0, 1, 2, 3, 4, 5], dtype=int64)

Model Building

```
In [23]: ## count vectorizer

count_vec = CountVectorizer(max_df=0.90,min_df=2,
                           max_features=1000,stop_words='english')

bagofword_vec = count_vec.fit_transform(books['summary'])
bagofword_vec
```

Out[23]: <3000x1000 sparse matrix of type '<class 'numpy.int64'>'
with 245372 stored elements in Compressed Sparse Row format>

```
In [24]: test = books['genre']
X_train, X_test, y_train, y_test = train_test_split(bagofword_vec,test,
                                                    test_size=0.2)

X_train.shape,X_test.shape
```

Out[24]: ((2400, 1000), (600, 1000))

```
In [25]: svc = SVC()
svc.fit(X_train,y_train)
svccpred = svc.predict(X_test)
print(metrics.accuracy_score(y_test,svccpred))

0.5483333333333333
```

```
In [26]: mb = MultinomialNB()
mb.fit(X_train,y_train)
mbpred = mb.predict(X_test)
print(metrics.accuracy_score(y_test,mbpred))
```

0.64

```
In [27]: rf = RandomForestClassifier()
rf.fit(X_train,y_train)
print(metrics.accuracy_score(y_test,rf.predict(X_test)))
```

0.6

Part 2 Model Building

- Changing from Countvectorizer to TFDIF vectorizer

```
In [28]: #Labeling each 'genre' with an unique number

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
y=le.fit_transform(books['genre'])

X_train,X_test,y_train,y_test = train_test_split(books['summary']
                                                ,y,test_size=0.2,
                                                random_state=557)

X_train.shape,X_test.shape
```

Out[28]: ((2400,), (600,))

```
In [29]: #Performing tf-idf

tfidf_vectorizer = TfidfVectorizer(max_df=0.8, max_features=10000)
xtrain_tfidf = tfidf_vectorizer.fit_transform(X_train.values.astype('U'))
xtest_tfidf = tfidf_vectorizer.transform(X_test.values.astype('U'))
```

```
In [30]: svc = SVC()
svc.fit(xtrain_tfidf,y_train)
svccpred = svc.predict(xtest_tfidf)
print(metrics.accuracy_score(y_test,svccpred))
```

0.7733333333333333

```
In [31]: mb = MultinomialNB()
mb.fit(xtrain_tfidf,y_train)
mbpred = mb.predict(xtest_tfidf)
print(metrics.accuracy_score(y_test,mbpred))
```

0.775

Testing the Model

In [32]:

```
def test(text,model):

    text = cleantext(text)
    text = removestopwords(text)
    text = lematizing(text)
    text = stemming(text)

    text_vector = tfidf_vectorizer.transform([text])
    predicted = model.predict(text_vector)
    return predicted

ans = books['summary'].apply(lambda text:test(text,mb))

# printing the
# print(list(mapper.keys())[list(mapper.values()).index(ans)])

ans
```

Out[32]:

```
0      [0]
1      [0]
2      [0]
3      [0]
4      [0]
...
2995   [1]
2996   [5]
2997   [5]
2998   [5]
2999   [5]
Name: summary, Length: 3000, dtype: object
```

In [33]:

```
predicted_genres = []
for i in range(len(ans)):

    index_val = ans[i][0]
    predicted_genres.append(list(mapper.keys())[list(mapper.values()).index(index_val)])
```

In [34]:

```
mapper
```

Out[34]:

```
{'Fantasy': 0,
 'Science Fiction': 1,
 'Crime Fiction': 2,
 'Historical novel': 3,
 'Horror': 4,
 'Thriller': 5}
```

In [35]:

```
## mapping the training genre as well

newmap = dict([(value,key) for key,value in mapper.items()])
newmap
```

Out[35]:

```
{0: 'Fantasy',
1: 'Science Fiction',
2: 'Crime Fiction',
3: 'Historical novel',
4: 'Horror',
5: 'Thriller'}
```

In [36]:

```
print(newmap)
```

```
{0: 'Fantasy', 1: 'Science Fiction', 2: 'Crime Fiction', 3: 'Historical novel', 4: 'Horror', 5: 'Thriller'}
```

In [37]:

```
books['Actual Genre'] = books['genre'].map(newmap)
books['Predicted_genre'] = np.array(predicted_genres)
books.head()
```

Out[37]:

	book_id	book_name	genre	summary	Actual Genre	Predicted_genre
0	3248537	Drowned Wednesday	0	drown wednesday first truste among morrow day ...	Fantasy	Fantasy
1	27796919	The Lost Hero	0	book open jason awaken school bu unabl rememb ...	Fantasy	Fantasy
2	3910776	The Eyes of the Overworld	0	cugel easili persuad merchant fianosth attempt...	Fantasy	Fantasy
3	5969644	Magic's Promise	0	book open herald mage vanyel return countri va...	Fantasy	Fantasy
4	3173445	Taran Wanderer	0	taran gurgi return caer dallben follow event t...	Fantasy	Fantasy

In [38]:

```
books = books[['book_name','summary','Actual Genre','Predicted_genre']]
books
```

Out[38]:

	book_name	summary	Actual Genre	Predicted_genre
0	Drowned Wednesday	drown wednesday first truste among morrow day ...	Fantasy	Fantasy
1	The Lost Hero	book open jason awaken school bu unabl rememb ...	Fantasy	Fantasy
2	The Eyes of the Overworld	cugel easili persuad merchant fianosth attempt...	Fantasy	Fantasy
3	Magic's Promise	book open herald mage vanyel return countri va...	Fantasy	Fantasy
4	Taran Wanderer	taran gurgi return caer dallben follow event t...	Fantasy	Fantasy
...
2995	White Death	novel numa file kurt austin adventur novel mai...	Thriller	Science Fiction
2996	Venus with Pistol	gilbert kemp dealer special antiqu gun london ...	Thriller	Thriller
2997	Blackwater	know deep davey alway live shadow older brothe...	Thriller	Thriller
2998	The Rainbow and the Rose	stori concern life johnni pasco retir commerci...	Thriller	Thriller
2999	Chiefs	first chief henri lee novel open grow town del...	Thriller	Thriller

3000 rows × 4 columns

In [39]: dict(Counter(books['Actual Genre'].values))

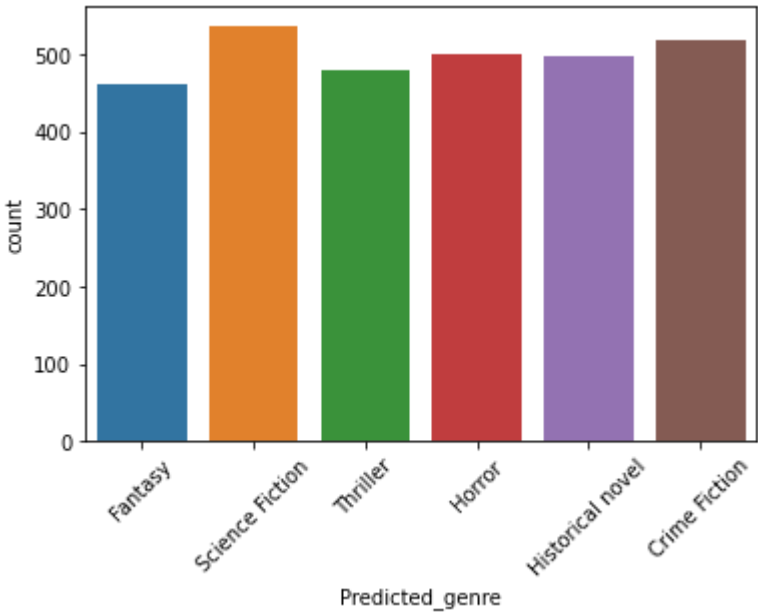
Out[39]: {'Fantasy': 500,
'Science Fiction': 500,
'Crime Fiction': 500,
'Historical novel': 500,
'Horror': 500,
'Thriller': 500}

In [40]: dict(Counter(books['Predicted_genre'].values))

Out[40]: {'Fantasy': 462,
'Science Fiction': 537,
'Thriller': 481,
'Horror': 502,
'Historical novel': 499,
'Crime Fiction': 519}

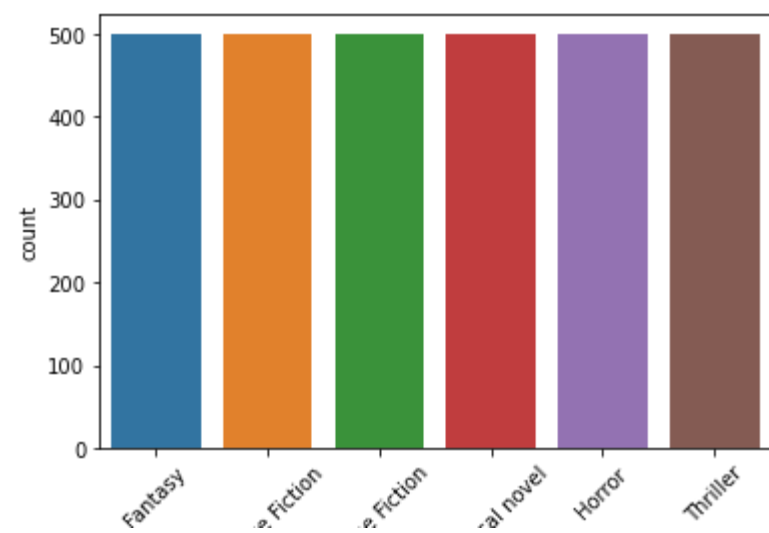
In [41]: sn.countplot(x = books['Predicted_genre'])
plt.xticks(rotation = '45')

Out[41]: (array([0, 1, 2, 3, 4, 5]),
[Text(0, 0, 'Fantasy'),
Text(1, 0, 'Science Fiction'),
Text(2, 0, 'Thriller'),
Text(3, 0, 'Horror'),
Text(4, 0, 'Historical novel'),
Text(5, 0, 'Crime Fiction')])



In [42]: sn.countplot(x = books['Actual Genre'])
plt.xticks(rotation = '45')

Out[42]: (array([0, 1, 2, 3, 4, 5]),
[Text(0, 0, 'Fantasy'),
Text(1, 0, 'Science Fiction'),
Text(2, 0, 'Crime Fiction'),
Text(3, 0, 'Historical novel'),
Text(4, 0, 'Horror'),
Text(5, 0, 'Thriller')])



```
In [43]: # saving the model

import pickle
file = open('bookgenremodel.pkl','wb')
pickle.dump(mb,file)
file.close()
```

```
In [44]: books['summary'].iloc[1]
```

Out[44]: 'book open jason awaken school bu unabl rememb anyth past sit next piper mclean leo valdez call name say girlfriend best friend respect three part class field trip grand canyon arriv classmat dylan turn venti sto
rm spirit attack trio trip leader coach gleeson hedg ensu fight jason surpris everyon includ one coin turn sword us battl storm spirit coach hedg reveal satyr fight taken captiv flee spirit battl fli chariot arri
v rescu trio one peopl annabeth upset discov miss boyfriend perci jackson expect annabeth seek perci told vision goddess hera look guy one shoe turn jason shoe destroy fight jason piper leo told demigod taken bac
k camp half blood meet greek demigod child like leo reveal son hephaestu piper daughter aphrodit jason son zeu though hera tell champion jason later discov full brother zeu demigod daughter thalia grace hunter ar
temi shortli arriv three given quest rescu hera captur set soon discov enemi work order gaea overthrow god quest encount thalia hunter look perci thalia jason reunite first sinc jason captur age two way aeolu cast
l jason leo piper becom separ thalia promis meet wolf hous last place thalia seen jason meet nearli apprehend aeolu gaea order trio manag escap thank melli aeolu former assist end san francisco thank result dream
piper aphrodit land san francisco trio rush mt diablo fight giant enceladu kidnap piper father manag kill giant save piper father rush wolf hous free hera although hero hunter save hera king giant porphyron rise
fulli disappear hole earth jason memori start return rememb hero roman counterpart camp half blood somewher near san francisco son jupit zeu roman aspect realiz hera also known juno switch perci jackson roman cam
p memori life hope two camp would ultim work togeth fight giant defeat goddess gaea'

```
In [45]: tfidf_vectorizer
```

Out[45]: TfidfVectorizer(max_df=0.8, max_features=10000)

```
In [46]: file = open('tfidfvector.pkl','wb')
pickle.dump(tfidf_vectorizer,file)
file.close()
```

```
In [47]: wordnet_down = nltk.download('wordnet',download_dir='./')
wordnet_down
```

[nltk_data] Downloading package wordnet to ./...
[nltk_data] Unzipping corpora\wordnet.zip.

Out[47]: True

```
In [48]: nltk.download('stopwords',download_dir='./')
```

[nltk_data] Downloading package stopwords to ./...
[nltk_data] Unzipping corpora\stopwords.zip.

Out[48]: True

