{

"nbformat": 4,

"nbformat\_minor": 0,

"metadata": {

"colab": {

"provenance": [],

"gpuType": "T4"

},

"kernelspec": {

"name": "python3",

"display\_name": "Python 3"

},

"language\_info": {

"name": "python"

},

"accelerator": "GPU"

},

"cells": [

{

"cell\_type": "markdown",

"source": [

"# Importing libraries"

],

"metadata": {

"id": "zEm10OYIJH\_Y"

}

},

{

"cell\_type": "markdown",

"source": [

"### These stereos download in certain packages and libraries for data analysis, for developing an application of machine learning, and for natural language processing. Some of additional packages are NumPy, Pandas, Matplotlib & Seaborn for data handling and visualization of data, TensorFlow & Keras for constructing a Neural Networks and NLTK for lemmatization & removal of stopwords. It also imports functions for constructing the train and test datasets; for assessing the accuracy of the model; and the tokenization. These warnings are turned of to facilitate going through a process without being interfered with."

],

"metadata": {

"id": "qYuT-ttpRnCV"

}

},

{

"cell\_type": "code",

"source": [

"import numpy as nmpy#The library named numpy is imported by using alias nmpy\n",

"import pandas as pnds#The library named pandas is imported by using alias pnds\n",

"import matplotlib.pyplot as mt\_plot#The library named matplotlib.pyplot is imported by using alias mt\_plot\n",

"import seaborn as sbrn # The library named seaborn is imported by using alias sbrn\n",

"from wordcloud import WordCloud # The library named wordcloud is imported\n",

"import nltk # The Natural Language Toolkit (nltk) is imported\n",

"nltk.download('stopwords') # The stopwords corpus is downloaded\n",

"nltk.download('punkt') # The punkt tokenizer is downloaded\n",

"nltk.download('wordnet') # The WordNet corpus is downloaded\n",

"from nltk.stem import WordNetLemmatizer # The WordNetLemmatizer class from nltk.stem is imported\n",

"from nltk.corpus import stopwords # The stopwords corpus from nltk.corpus is imported\n",

"from nltk.stem.porter import PorterStemmer # The Porter Stemmer is imported from nltk.stem.porter\n",

"import re # The regular expression library is imported\n",

"import tensorflow as tf # The TensorFlow library is imported\n",

"from keras.models import Model # The Model class from keras.models is imported\n",

"from keras.layers import Input, Embedding, LSTM, Dense, Flatten, Reshape # Specific layers from keras.layers are imported\n",

"from keras.optimizers import Adam # The Adam optimizer is imported from keras.optimizers\n",

"from keras.metrics import Precision, Recall # Precision and Recall metrics from keras.metrics are imported\n",

"from sklearn.model\_selection import train\_test\_split # The train\_test\_split function from sklearn.model\_selection is imported\n",

"from sklearn.metrics import confusion\_matrix, accuracy\_score, precision\_score, recall\_score, f1\_score, classification\_report\n",

"from tensorflow.keras.preprocessing.text import Tokenizer # The Tokenizer class from tensorflow.keras.preprocessing.text is imported\n",

"from tensorflow.keras.preprocessing.sequence import pad\_sequences # The pad\_sequences function from tensorflow.keras.preprocessing.sequence is imported\n",

"\n",

"import warnings # The warnings module is imported\n",

"warnings.filterwarnings('ignore') # All warnings are ignored"

],

"metadata": {

"id": "DPPvOAZBRca9"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"# Unzip data in drive in the same location\n",

"!unzip '/content/drive/MyDrive/Colab Notebooks/Anjaneyam NLP Code /archive (14).zip' -d '/content/drive/MyDrive/Colab Notebooks/Anjaneyam NLP Code /'"

],

"metadata": {

"id": "wS33oQjI0vGy"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"imdb\_data = pnds.read\_csv('/content/drive/MyDrive/Colab Notebooks/Anjaneyam NLP Code /IMDB Dataset.csv')"

],

"metadata": {

"id": "LT1wItlW7OJk"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"imdb\_data.head()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 206

},

"id": "66hV9SNz7UpE",

"outputId": "1c866c68-0fb1-430d-bbf9-13d9a07ed4b8"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" review sentiment\n",

"0 One of the other reviewers has mentioned that ... positive\n",

"1 A wonderful little production. <br /><br />The... positive\n",

"2 I thought this was a wonderful way to spend ti... positive\n",

"3 Basically there's a family where a little boy ... negative\n",

"4 Petter Mattei's \"Love in the Time of Money\" is... positive"

],

"text/html": [

"\n",

" <div id=\"df-d40e5684-4bca-4881-900a-9735f75a3ecc\" class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>review</th>\n",

" <th>sentiment</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>One of the other reviewers has mentioned that ...</td>\n",

" <td>positive</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>A wonderful little production. &lt;br /&gt;&lt;br /&gt;The...</td>\n",

" <td>positive</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>I thought this was a wonderful way to spend ti...</td>\n",

" <td>positive</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>Basically there's a family where a little boy ...</td>\n",

" <td>negative</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>Petter Mattei's \"Love in the Time of Money\" is...</td>\n",

" <td>positive</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>\n",

" <div class=\"colab-df-buttons\">\n",

"\n",

" <div class=\"colab-df-container\">\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-d40e5684-4bca-4881-900a-9735f75a3ecc')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

"\n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\" viewBox=\"0 -960 960 960\">\n",

" <path d=\"M120-120v-720h720v720H120Zm60-500h600v-160H180v160Zm220 220h160v-160H400v160Zm0 220h160v-160H400v160ZM180-400h160v-160H180v160Zm440 0h160v-160H620v160ZM180-180h160v-160H180v160Zm440 0h160v-160H620v160Z\"/>\n",

" </svg>\n",

" </button>\n",

"\n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" .colab-df-buttons div {\n",

" margin-bottom: 4px;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-d40e5684-4bca-4881-900a-9735f75a3ecc button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-d40e5684-4bca-4881-900a-9735f75a3ecc');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

"\n",

"\n",

"<div id=\"df-92f7d067-4923-4106-bcdd-2ff55aba34d2\">\n",

" <button class=\"colab-df-quickchart\" onclick=\"quickchart('df-92f7d067-4923-4106-bcdd-2ff55aba34d2')\"\n",

" title=\"Suggest charts\"\n",

" style=\"display:none;\">\n",

"\n",

"<svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <g>\n",

" <path d=\"M19 3H5c-1.1 0-2 .9-2 2v14c0 1.1.9 2 2 2h14c1.1 0 2-.9 2-2V5c0-1.1-.9-2-2-2zM9 17H7v-7h2v7zm4 0h-2V7h2v10zm4 0h-2v-4h2v4z\"/>\n",

" </g>\n",

"</svg>\n",

" </button>\n",

"\n",

"<style>\n",

" .colab-df-quickchart {\n",

" --bg-color: #E8F0FE;\n",

" --fill-color: #1967D2;\n",

" --hover-bg-color: #E2EBFA;\n",

" --hover-fill-color: #174EA6;\n",

" --disabled-fill-color: #AAA;\n",

" --disabled-bg-color: #DDD;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-quickchart {\n",

" --bg-color: #3B4455;\n",

" --fill-color: #D2E3FC;\n",

" --hover-bg-color: #434B5C;\n",

" --hover-fill-color: #FFFFFF;\n",

" --disabled-bg-color: #3B4455;\n",

" --disabled-fill-color: #666;\n",

" }\n",

"\n",

" .colab-df-quickchart {\n",

" background-color: var(--bg-color);\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: var(--fill-color);\n",

" height: 32px;\n",

" padding: 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-quickchart:hover {\n",

" background-color: var(--hover-bg-color);\n",

" box-shadow: 0 1px 2px rgba(60, 64, 67, 0.3), 0 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: var(--button-hover-fill-color);\n",

" }\n",

"\n",

" .colab-df-quickchart-complete:disabled,\n",

" .colab-df-quickchart-complete:disabled:hover {\n",

" background-color: var(--disabled-bg-color);\n",

" fill: var(--disabled-fill-color);\n",

" box-shadow: none;\n",

" }\n",

"\n",

" .colab-df-spinner {\n",

" border: 2px solid var(--fill-color);\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" animation:\n",

" spin 1s steps(1) infinite;\n",

" }\n",

"\n",

" @keyframes spin {\n",

" 0% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" border-left-color: var(--fill-color);\n",

" }\n",

" 20% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 30% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 40% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 60% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 80% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" 90% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" }\n",

"</style>\n",

"\n",

" <script>\n",

" async function quickchart(key) {\n",

" const quickchartButtonEl =\n",

" document.querySelector('#' + key + ' button');\n",

" quickchartButtonEl.disabled = true; // To prevent multiple clicks.\n",

" quickchartButtonEl.classList.add('colab-df-spinner');\n",

" try {\n",

" const charts = await google.colab.kernel.invokeFunction(\n",

" 'suggestCharts', [key], {});\n",

" } catch (error) {\n",

" console.error('Error during call to suggestCharts:', error);\n",

" }\n",

" quickchartButtonEl.classList.remove('colab-df-spinner');\n",

" quickchartButtonEl.classList.add('colab-df-quickchart-complete');\n",

" }\n",

" (() => {\n",

" let quickchartButtonEl =\n",

" document.querySelector('#df-92f7d067-4923-4106-bcdd-2ff55aba34d2 button');\n",

" quickchartButtonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

" })();\n",

" </script>\n",

"</div>\n",

"\n",

" </div>\n",

" </div>\n"

],

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "dataframe",

"variable\_name": "imdb\_data",

"summary": "{\n \"name\": \"imdb\_data\",\n \"rows\": 50000,\n \"fields\": [\n {\n \"column\": \"review\",\n \"properties\": {\n \"dtype\": \"string\",\n \"num\_unique\_values\": 49582,\n \"samples\": [\n \"\\\"Soul Plane\\\" is a horrible attempt at comedy that only should appeal people with thick skulls, bloodshot eyes and furry pawns. <br /><br />The plot is not only incoherent but also non-existent, acting is mostly sub sub-par with a gang of highly moronic and dreadful characters thrown in for bad measure, jokes are often spotted miles ahead and almost never even a bit amusing. This movie lacks any structure and is full of racial stereotypes that must have seemed old even in the fifties, the only thing it really has going for it is some pretty ladies, but really, if you want that you can rent something from the \\\"Adult\\\" section. OK?<br /><br />I can hardly see anything here to recommend since you'll probably have a lot a better and productive time chasing rats with a sledgehammer or inventing waterproof teabags or whatever.<br /><br />2/10\",\n \"Guest from the Future tells a fascinating story of time travel, friendship, battle of good and evil -- all with a small budget, child actors, and few special effects. Something for Spielberg and Lucas to learn from. ;) A sixth-grader Kolya \\\"Nick\\\" Gerasimov finds a time machine in the basement of a decrepit building and travels 100 years into the future. He discovers a near-perfect, utopian society where robots play guitars and write poetry, everyone is kind to each other and people enjoy everything technology has to offer. Alice is the daughter of a prominent scientist who invented a device called Mielophone that allows to read minds of humans and animals. The device can be put to both good and bad use, depending on whose hands it falls into. When two evil space pirates from Saturn who want to rule the universe attempt to steal Mielophone, it falls into the hands of 20th century school boy Nick. With the pirates hot on his tracks, he travels back to his time, followed by the pirates, and Alice. Chaos, confusion and funny situations follow as the luckless pirates try to blend in with the earthlings. Alice enrolls in the same school Nick goes to and demonstrates superhuman abilities in PE class. The catch is, Alice doesn't know what Nick looks like, while the pirates do. Also, the pirates are able to change their appearance and turn literally into anyone. (Hmm, I wonder if this is where James Cameron got the idea for Terminator...) Who gets to Nick -- and Mielophone -- first? Excellent plot, non-stop adventures, and great soundtrack. I wish Hollywood made kid movies like this one...\",\n \"\\\"National Treasure\\\" (2004) is a thoroughly misguided hodge-podge of plot entanglements that borrow from nearly every cloak and dagger government conspiracy clich\\u00e9 that has ever been written. The film stars Nicholas Cage as Benjamin Franklin Gates (how precious is that, I ask you?); a seemingly normal fellow who, for no other reason than being of a lineage of like-minded misguided fortune hunters, decides to steal a 'national treasure' that has been hidden by the United States founding fathers. After a bit of subtext and background that plays laughably (unintentionally) like Indiana Jones meets The Patriot, the film degenerates into one misguided whimsy after another \\u0096 attempting to create a 'Stanley Goodspeed' regurgitation of Nicholas Cage and launch the whole convoluted mess forward with a series of high octane, but disconnected misadventures.<br /><br />The relevancy and logic to having George Washington and his motley crew of patriots burying a king's ransom someplace on native soil, and then, going through the meticulous plan of leaving clues scattered throughout U.S. currency art work, is something that director Jon Turteltaub never quite gets around to explaining. Couldn't Washington found better usage for such wealth during the start up of the country? Hence, we are left with a mystery built on top of an enigma that is already on shaky ground by the time Ben appoints himself the new custodian of this untold wealth. Ben's intentions are noble \\u0096 if confusing. He's set on protecting the treasure. For who and when?\\u0085your guess is as good as mine.<br /><br />But there are a few problems with Ben's crusade. First up, his friend, Ian Holmes (Sean Bean) decides that he can't wait for Ben to make up his mind about stealing the Declaration of Independence from the National Archives (oh, yeah \\u0096 brilliant idea!). Presumably, the back of that famous document holds the secret answer to the ultimate fortune. So Ian tries to kill Ben. The assassination attempt is, of course, unsuccessful, if overly melodramatic. It also affords Ben the opportunity to pick up, and pick on, the very sultry curator of the archives, Abigail Chase (Diane Kruger). She thinks Ben is clearly a nut \\u0096 at least at the beginning. But true to action/romance form, Abby's resolve melts quicker than you can say, \\\"is that the Hope Diamond?\\\" The film moves into full X-File-ish mode, as the FBI, mistakenly believing that Ben is behind the theft, retaliate in various benign ways that lead to a multi-layering of action sequences reminiscent of Mission Impossible meets The Fugitive. Honestly, don't those guys ever get 'intelligence' information that is correct? In the final analysis, \\\"National Treasure\\\" isn't great film making, so much as it's a patchwork rehash of tired old bits from other movies, woven together from scraps, the likes of which would make IL' Betsy Ross blush.<br /><br />The Buena Vista DVD delivers a far more generous treatment than this film is deserving of. The anamorphic widescreen picture exhibits a very smooth and finely detailed image with very rich colors, natural flesh tones, solid blacks and clean whites. The stylized image is also free of blemishes and digital enhancements. The audio is 5.1 and delivers a nice sonic boom to your side and rear speakers with intensity and realism. Extras include a host of promotional junket material that is rather deep and over the top in its explanation of how and why this film was made. If only, as an audience, we had had more clarification as to why Ben and co. were chasing after an illusive treasure, this might have been one good flick. Extras conclude with the theatrical trailer, audio commentary and deleted scenes. Not for the faint-hearted \\u0096 just the thick-headed.\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n },\n {\n \"column\": \"sentiment\",\n \"properties\": {\n \"dtype\": \"category\",\n \"num\_unique\_values\": 2,\n \"samples\": [\n \"negative\",\n \"positive\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n }\n ]\n}"

}

},

"metadata": {},

"execution\_count": 44

}

]

},

{

"cell\_type": "code",

"source": [

"imdb\_data.shape"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "MzH9urLF7Wgn",

"outputId": "c68f4977-abae-4b2c-c190-ff9ddbe1a6e6"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"(50000, 2)"

]

},

"metadata": {},

"execution\_count": 45

}

]

},

{

"cell\_type": "code",

"source": [

"imdb\_data.sentiment.value\_counts()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 141

},

"id": "7lcyhe8n7YWC",

"outputId": "4b1f5001-c68f-4097-d06c-bbff15d0d2e6",

"collapsed": true

},

"execution\_count": null,

"outputs": [

{

"output\_type": "error",

"ename": "NameError",

"evalue": "name 'imdb\_data' is not defined",

"traceback": [

"\u001b[0;31m---------------------------------------------------------------------------\u001b[0m",

"\u001b[0;31mNameError\u001b[0m Traceback (most recent call last)",

"\u001b[0;32m<ipython-input-1-e731907d067d>\u001b[0m in \u001b[0;36m<cell line: 1>\u001b[0;34m()\u001b[0m\n\u001b[0;32m----> 1\u001b[0;31m \u001b[0mimdb\_data\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0msentiment\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mvalue\_counts\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m",

"\u001b[0;31mNameError\u001b[0m: name 'imdb\_data' is not defined"

]

}

]

},

{

"cell\_type": "code",

"source": [

"imdb\_data.isnull().sum()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 146

},

"id": "65\_6chra7Z6o",

"outputId": "29ad057f-594e-4d77-fecd-14e7c3644e34"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"review 0\n",

"sentiment 0\n",

"dtype: int64"

],

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>0</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>review</th>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>sentiment</th>\n",

" <td>0</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div><br><label><b>dtype:</b> int64</label>"

]

},

"metadata": {},

"execution\_count": 47

}

]

},

{

"cell\_type": "code",

"source": [

"imdb\_data.duplicated().sum()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "SOmc6Ll27sKh",

"outputId": "7843567a-14e1-4ffc-98b7-b6d464d62fc4"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"418"

]

},

"metadata": {},

"execution\_count": 48

}

]

},

{

"cell\_type": "code",

"source": [

"imdb\_data.drop\_duplicates(inplace=True)"

],

"metadata": {

"id": "3yxve9zX7uUq"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"imdb\_data.duplicated().sum()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "ySzPv14b7wKp",

"outputId": "e584c555-fdf1-41c6-d97a-1e6ca32d38cc"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"0"

]

},

"metadata": {},

"execution\_count": 50

}

]

},

{

"cell\_type": "code",

"source": [

"imdb\_data.sentiment.value\_counts()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 178

},

"id": "TFFpzR0y7yAt",

"outputId": "72836405-88d6-43cf-fe94-3c9964e76abe"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"sentiment\n",

"positive 24884\n",

"negative 24698\n",

"Name: count, dtype: int64"

],

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>count</th>\n",

" </tr>\n",

" <tr>\n",

" <th>sentiment</th>\n",

" <th></th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>positive</th>\n",

" <td>24884</td>\n",

" </tr>\n",

" <tr>\n",

" <th>negative</th>\n",

" <td>24698</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div><br><label><b>dtype:</b> int64</label>"

]

},

"metadata": {},

"execution\_count": 51

}

]

},

{

"cell\_type": "code",

"source": [

"imdb\_data['sentiment'] = imdb\_data['sentiment'].replace({'psitiv': 1, 'negative': 0})\n"

],

"metadata": {

"id": "1lc\_cgLj70M0"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"imdb\_data.head()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 206

},

"id": "jGWWegjj8sk3",

"outputId": "a78d4f87-067e-41ec-e920-be558d71b80c"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" review sentiment\n",

"0 One of the other reviewers has mentioned that ... 1\n",

"1 A wonderful little production. <br /><br />The... 1\n",

"2 I thought this was a wonderful way to spend ti... 1\n",

"3 Basically there's a family where a little boy ... 0\n",

"4 Petter Mattei's \"Love in the Time of Money\" is... 1"

],

"text/html": [

"\n",

" <div id=\"df-0c5ceed2-adae-42b5-92d6-85459bce9429\" class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>review</th>\n",

" <th>sentiment</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>One of the other reviewers has mentioned that ...</td>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>A wonderful little production. &lt;br /&gt;&lt;br /&gt;The...</td>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>I thought this was a wonderful way to spend ti...</td>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>Basically there's a family where a little boy ...</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>Petter Mattei's \"Love in the Time of Money\" is...</td>\n",

" <td>1</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>\n",

" <div class=\"colab-df-buttons\">\n",

"\n",

" <div class=\"colab-df-container\">\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-0c5ceed2-adae-42b5-92d6-85459bce9429')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

"\n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\" viewBox=\"0 -960 960 960\">\n",

" <path d=\"M120-120v-720h720v720H120Zm60-500h600v-160H180v160Zm220 220h160v-160H400v160Zm0 220h160v-160H400v160ZM180-400h160v-160H180v160Zm440 0h160v-160H620v160ZM180-180h160v-160H180v160Zm440 0h160v-160H620v160Z\"/>\n",

" </svg>\n",

" </button>\n",

"\n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" .colab-df-buttons div {\n",

" margin-bottom: 4px;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-0c5ceed2-adae-42b5-92d6-85459bce9429 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-0c5ceed2-adae-42b5-92d6-85459bce9429');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

"\n",

"\n",

"<div id=\"df-e41b3cdd-e297-4a4b-972e-222fb5900b57\">\n",

" <button class=\"colab-df-quickchart\" onclick=\"quickchart('df-e41b3cdd-e297-4a4b-972e-222fb5900b57')\"\n",

" title=\"Suggest charts\"\n",

" style=\"display:none;\">\n",

"\n",

"<svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <g>\n",

" <path d=\"M19 3H5c-1.1 0-2 .9-2 2v14c0 1.1.9 2 2 2h14c1.1 0 2-.9 2-2V5c0-1.1-.9-2-2-2zM9 17H7v-7h2v7zm4 0h-2V7h2v10zm4 0h-2v-4h2v4z\"/>\n",

" </g>\n",

"</svg>\n",

" </button>\n",

"\n",

"<style>\n",

" .colab-df-quickchart {\n",

" --bg-color: #E8F0FE;\n",

" --fill-color: #1967D2;\n",

" --hover-bg-color: #E2EBFA;\n",

" --hover-fill-color: #174EA6;\n",

" --disabled-fill-color: #AAA;\n",

" --disabled-bg-color: #DDD;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-quickchart {\n",

" --bg-color: #3B4455;\n",

" --fill-color: #D2E3FC;\n",

" --hover-bg-color: #434B5C;\n",

" --hover-fill-color: #FFFFFF;\n",

" --disabled-bg-color: #3B4455;\n",

" --disabled-fill-color: #666;\n",

" }\n",

"\n",

" .colab-df-quickchart {\n",

" background-color: var(--bg-color);\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: var(--fill-color);\n",

" height: 32px;\n",

" padding: 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-quickchart:hover {\n",

" background-color: var(--hover-bg-color);\n",

" box-shadow: 0 1px 2px rgba(60, 64, 67, 0.3), 0 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: var(--button-hover-fill-color);\n",

" }\n",

"\n",

" .colab-df-quickchart-complete:disabled,\n",

" .colab-df-quickchart-complete:disabled:hover {\n",

" background-color: var(--disabled-bg-color);\n",

" fill: var(--disabled-fill-color);\n",

" box-shadow: none;\n",

" }\n",

"\n",

" .colab-df-spinner {\n",

" border: 2px solid var(--fill-color);\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" animation:\n",

" spin 1s steps(1) infinite;\n",

" }\n",

"\n",

" @keyframes spin {\n",

" 0% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" border-left-color: var(--fill-color);\n",

" }\n",

" 20% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 30% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 40% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 60% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 80% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" 90% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" }\n",

"</style>\n",

"\n",

" <script>\n",

" async function quickchart(key) {\n",

" const quickchartButtonEl =\n",

" document.querySelector('#' + key + ' button');\n",

" quickchartButtonEl.disabled = true; // To prevent multiple clicks.\n",

" quickchartButtonEl.classList.add('colab-df-spinner');\n",

" try {\n",

" const charts = await google.colab.kernel.invokeFunction(\n",

" 'suggestCharts', [key], {});\n",

" } catch (error) {\n",

" console.error('Error during call to suggestCharts:', error);\n",

" }\n",

" quickchartButtonEl.classList.remove('colab-df-spinner');\n",

" quickchartButtonEl.classList.add('colab-df-quickchart-complete');\n",

" }\n",

" (() => {\n",

" let quickchartButtonEl =\n",

" document.querySelector('#df-e41b3cdd-e297-4a4b-972e-222fb5900b57 button');\n",

" quickchartButtonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

" })();\n",

" </script>\n",

"</div>\n",

"\n",

" </div>\n",

" </div>\n"

],

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "dataframe",

"variable\_name": "imdb\_data",

"summary": "{\n \"name\": \"imdb\_data\",\n \"rows\": 49582,\n \"fields\": [\n {\n \"column\": \"review\",\n \"properties\": {\n \"dtype\": \"string\",\n \"num\_unique\_values\": 49582,\n \"samples\": [\n \"\\\"Soul Plane\\\" is a horrible attempt at comedy that only should appeal people with thick skulls, bloodshot eyes and furry pawns. <br /><br />The plot is not only incoherent but also non-existent, acting is mostly sub sub-par with a gang of highly moronic and dreadful characters thrown in for bad measure, jokes are often spotted miles ahead and almost never even a bit amusing. This movie lacks any structure and is full of racial stereotypes that must have seemed old even in the fifties, the only thing it really has going for it is some pretty ladies, but really, if you want that you can rent something from the \\\"Adult\\\" section. OK?<br /><br />I can hardly see anything here to recommend since you'll probably have a lot a better and productive time chasing rats with a sledgehammer or inventing waterproof teabags or whatever.<br /><br />2/10\",\n \"Guest from the Future tells a fascinating story of time travel, friendship, battle of good and evil -- all with a small budget, child actors, and few special effects. Something for Spielberg and Lucas to learn from. ;) A sixth-grader Kolya \\\"Nick\\\" Gerasimov finds a time machine in the basement of a decrepit building and travels 100 years into the future. He discovers a near-perfect, utopian society where robots play guitars and write poetry, everyone is kind to each other and people enjoy everything technology has to offer. Alice is the daughter of a prominent scientist who invented a device called Mielophone that allows to read minds of humans and animals. The device can be put to both good and bad use, depending on whose hands it falls into. When two evil space pirates from Saturn who want to rule the universe attempt to steal Mielophone, it falls into the hands of 20th century school boy Nick. With the pirates hot on his tracks, he travels back to his time, followed by the pirates, and Alice. Chaos, confusion and funny situations follow as the luckless pirates try to blend in with the earthlings. Alice enrolls in the same school Nick goes to and demonstrates superhuman abilities in PE class. The catch is, Alice doesn't know what Nick looks like, while the pirates do. Also, the pirates are able to change their appearance and turn literally into anyone. (Hmm, I wonder if this is where James Cameron got the idea for Terminator...) Who gets to Nick -- and Mielophone -- first? Excellent plot, non-stop adventures, and great soundtrack. I wish Hollywood made kid movies like this one...\",\n \"\\\"National Treasure\\\" (2004) is a thoroughly misguided hodge-podge of plot entanglements that borrow from nearly every cloak and dagger government conspiracy clich\\u00e9 that has ever been written. The film stars Nicholas Cage as Benjamin Franklin Gates (how precious is that, I ask you?); a seemingly normal fellow who, for no other reason than being of a lineage of like-minded misguided fortune hunters, decides to steal a 'national treasure' that has been hidden by the United States founding fathers. After a bit of subtext and background that plays laughably (unintentionally) like Indiana Jones meets The Patriot, the film degenerates into one misguided whimsy after another \\u0096 attempting to create a 'Stanley Goodspeed' regurgitation of Nicholas Cage and launch the whole convoluted mess forward with a series of high octane, but disconnected misadventures.<br /><br />The relevancy and logic to having George Washington and his motley crew of patriots burying a king's ransom someplace on native soil, and then, going through the meticulous plan of leaving clues scattered throughout U.S. currency art work, is something that director Jon Turteltaub never quite gets around to explaining. Couldn't Washington found better usage for such wealth during the start up of the country? Hence, we are left with a mystery built on top of an enigma that is already on shaky ground by the time Ben appoints himself the new custodian of this untold wealth. Ben's intentions are noble \\u0096 if confusing. He's set on protecting the treasure. For who and when?\\u0085your guess is as good as mine.<br /><br />But there are a few problems with Ben's crusade. First up, his friend, Ian Holmes (Sean Bean) decides that he can't wait for Ben to make up his mind about stealing the Declaration of Independence from the National Archives (oh, yeah \\u0096 brilliant idea!). Presumably, the back of that famous document holds the secret answer to the ultimate fortune. So Ian tries to kill Ben. The assassination attempt is, of course, unsuccessful, if overly melodramatic. It also affords Ben the opportunity to pick up, and pick on, the very sultry curator of the archives, Abigail Chase (Diane Kruger). She thinks Ben is clearly a nut \\u0096 at least at the beginning. But true to action/romance form, Abby's resolve melts quicker than you can say, \\\"is that the Hope Diamond?\\\" The film moves into full X-File-ish mode, as the FBI, mistakenly believing that Ben is behind the theft, retaliate in various benign ways that lead to a multi-layering of action sequences reminiscent of Mission Impossible meets The Fugitive. Honestly, don't those guys ever get 'intelligence' information that is correct? In the final analysis, \\\"National Treasure\\\" isn't great film making, so much as it's a patchwork rehash of tired old bits from other movies, woven together from scraps, the likes of which would make IL' Betsy Ross blush.<br /><br />The Buena Vista DVD delivers a far more generous treatment than this film is deserving of. The anamorphic widescreen picture exhibits a very smooth and finely detailed image with very rich colors, natural flesh tones, solid blacks and clean whites. The stylized image is also free of blemishes and digital enhancements. The audio is 5.1 and delivers a nice sonic boom to your side and rear speakers with intensity and realism. Extras include a host of promotional junket material that is rather deep and over the top in its explanation of how and why this film was made. If only, as an audience, we had had more clarification as to why Ben and co. were chasing after an illusive treasure, this might have been one good flick. Extras conclude with the theatrical trailer, audio commentary and deleted scenes. Not for the faint-hearted \\u0096 just the thick-headed.\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n },\n {\n \"column\": \"sentiment\",\n \"properties\": {\n \"dtype\": \"number\",\n \"std\": 0,\n \"min\": 0,\n \"max\": 1,\n \"num\_unique\_values\": 2,\n \"samples\": [\n 0,\n 1\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n }\n ]\n}"

}

},

"metadata": {},

"execution\_count": 53

}

]

},

{

"cell\_type": "markdown",

"source": [

"# Preprocessing"

],

"metadata": {

"id": "ByfQXUeB3Kt4"

}

},

{

"cell\_type": "markdown",

"source": [

"This code defines a function clean\_review to preprocess movie reviews by removing unwanted elements. It eliminates HTML tags, URLs, and non-alphabetic characters from the text, converts it to lowercase, and splits it into individual words before joining them back into a cleaned string. The function is then applied to a dataset imdb\_data containing movie reviews, creating a new column called Processed\_review that holds the cleaned version of each review. The head() function displays the first few rows of the modified dataset."

],

"metadata": {

"id": "9JzZORl9Um6z"

}

},

{

"cell\_type": "code",

"source": [

"stop\_words = set(stopwords.words('english'))\n",

"\n",

"def clean\_review(Procesd\_Colmn):\n",

" Procesd\_Colmn = re.sub('<.\*?>', '', Procesd\_Colmn) # Remove HTML tags\n",

" Procesd\_Colmn = re.sub(r'http\\S+', '', Procesd\_Colmn) # Remove URLs\n",

" Procesd\_Colmn = re.sub('[^a-zA-Z]', ' ', Procesd\_Colmn) # Remove non-alphabetic characters\n",

" Procesd\_Colmn = Procesd\_Colmn.lower()\n",

" Procesd\_Colmn = Procesd\_Colmn.split()\n",

" Procesd\_Colmn = ' '.join(Procesd\_Colmn)\n",

" return Procesd\_Colmn\n",

"\n",

"\n",

"imdb\_data['Processed\_review'] = imdb\_data['review'].apply(clean\_review)\n",

"\n",

"imdb\_data.head()\n"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 278

},

"id": "-3w5VZgWAJr1",

"outputId": "e095acd9-77ac-4f96-c11e-c5a23020a1c8"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"[nltk\_data] Downloading package stopwords to /root/nltk\_data...\n",

"[nltk\_data] Package stopwords is already up-to-date!\n",

"[nltk\_data] Downloading package punkt to /root/nltk\_data...\n",

"[nltk\_data] Package punkt is already up-to-date!\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" review sentiment \\\n",

"0 One of the other reviewers has mentioned that ... 1 \n",

"1 A wonderful little production. <br /><br />The... 1 \n",

"2 I thought this was a wonderful way to spend ti... 1 \n",

"3 Basically there's a family where a little boy ... 0 \n",

"4 Petter Mattei's \"Love in the Time of Money\" is... 1 \n",

"\n",

" Processed\_review \n",

"0 one of the other reviewers has mentioned that ... \n",

"1 a wonderful little production the filming tech... \n",

"2 i thought this was a wonderful way to spend ti... \n",

"3 basically there s a family where a little boy ... \n",

"4 petter mattei s love in the time of money is a... "

],

"text/html": [

"\n",

" <div id=\"df-379d81ae-7a58-4d01-adad-e8f31871c78b\" class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>review</th>\n",

" <th>sentiment</th>\n",

" <th>Processed\_review</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>One of the other reviewers has mentioned that ...</td>\n",

" <td>1</td>\n",

" <td>one of the other reviewers has mentioned that ...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>A wonderful little production. &lt;br /&gt;&lt;br /&gt;The...</td>\n",

" <td>1</td>\n",

" <td>a wonderful little production the filming tech...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>I thought this was a wonderful way to spend ti...</td>\n",

" <td>1</td>\n",

" <td>i thought this was a wonderful way to spend ti...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>Basically there's a family where a little boy ...</td>\n",

" <td>0</td>\n",

" <td>basically there s a family where a little boy ...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>Petter Mattei's \"Love in the Time of Money\" is...</td>\n",

" <td>1</td>\n",

" <td>petter mattei s love in the time of money is a...</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>\n",

" <div class=\"colab-df-buttons\">\n",

"\n",

" <div class=\"colab-df-container\">\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-379d81ae-7a58-4d01-adad-e8f31871c78b')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

"\n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\" viewBox=\"0 -960 960 960\">\n",

" <path d=\"M120-120v-720h720v720H120Zm60-500h600v-160H180v160Zm220 220h160v-160H400v160Zm0 220h160v-160H400v160ZM180-400h160v-160H180v160Zm440 0h160v-160H620v160ZM180-180h160v-160H180v160Zm440 0h160v-160H620v160Z\"/>\n",

" </svg>\n",

" </button>\n",

"\n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" .colab-df-buttons div {\n",

" margin-bottom: 4px;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-379d81ae-7a58-4d01-adad-e8f31871c78b button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-379d81ae-7a58-4d01-adad-e8f31871c78b');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

"\n",

"\n",

"<div id=\"df-482f69d6-1a2a-4e7e-8e91-52007f75ab21\">\n",

" <button class=\"colab-df-quickchart\" onclick=\"quickchart('df-482f69d6-1a2a-4e7e-8e91-52007f75ab21')\"\n",

" title=\"Suggest charts\"\n",

" style=\"display:none;\">\n",

"\n",

"<svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <g>\n",

" <path d=\"M19 3H5c-1.1 0-2 .9-2 2v14c0 1.1.9 2 2 2h14c1.1 0 2-.9 2-2V5c0-1.1-.9-2-2-2zM9 17H7v-7h2v7zm4 0h-2V7h2v10zm4 0h-2v-4h2v4z\"/>\n",

" </g>\n",

"</svg>\n",

" </button>\n",

"\n",

"<style>\n",

" .colab-df-quickchart {\n",

" --bg-color: #E8F0FE;\n",

" --fill-color: #1967D2;\n",

" --hover-bg-color: #E2EBFA;\n",

" --hover-fill-color: #174EA6;\n",

" --disabled-fill-color: #AAA;\n",

" --disabled-bg-color: #DDD;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-quickchart {\n",

" --bg-color: #3B4455;\n",

" --fill-color: #D2E3FC;\n",

" --hover-bg-color: #434B5C;\n",

" --hover-fill-color: #FFFFFF;\n",

" --disabled-bg-color: #3B4455;\n",

" --disabled-fill-color: #666;\n",

" }\n",

"\n",

" .colab-df-quickchart {\n",

" background-color: var(--bg-color);\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: var(--fill-color);\n",

" height: 32px;\n",

" padding: 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-quickchart:hover {\n",

" background-color: var(--hover-bg-color);\n",

" box-shadow: 0 1px 2px rgba(60, 64, 67, 0.3), 0 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: var(--button-hover-fill-color);\n",

" }\n",

"\n",

" .colab-df-quickchart-complete:disabled,\n",

" .colab-df-quickchart-complete:disabled:hover {\n",

" background-color: var(--disabled-bg-color);\n",

" fill: var(--disabled-fill-color);\n",

" box-shadow: none;\n",

" }\n",

"\n",

" .colab-df-spinner {\n",

" border: 2px solid var(--fill-color);\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" animation:\n",

" spin 1s steps(1) infinite;\n",

" }\n",

"\n",

" @keyframes spin {\n",

" 0% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" border-left-color: var(--fill-color);\n",

" }\n",

" 20% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 30% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 40% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 60% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 80% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" 90% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" }\n",

"</style>\n",

"\n",

" <script>\n",

" async function quickchart(key) {\n",

" const quickchartButtonEl =\n",

" document.querySelector('#' + key + ' button');\n",

" quickchartButtonEl.disabled = true; // To prevent multiple clicks.\n",

" quickchartButtonEl.classList.add('colab-df-spinner');\n",

" try {\n",

" const charts = await google.colab.kernel.invokeFunction(\n",

" 'suggestCharts', [key], {});\n",

" } catch (error) {\n",

" console.error('Error during call to suggestCharts:', error);\n",

" }\n",

" quickchartButtonEl.classList.remove('colab-df-spinner');\n",

" quickchartButtonEl.classList.add('colab-df-quickchart-complete');\n",

" }\n",

" (() => {\n",

" let quickchartButtonEl =\n",

" document.querySelector('#df-482f69d6-1a2a-4e7e-8e91-52007f75ab21 button');\n",

" quickchartButtonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

" })();\n",

" </script>\n",

"</div>\n",

"\n",

" </div>\n",

" </div>\n"

],

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "dataframe",

"variable\_name": "imdb\_data",

"summary": "{\n \"name\": \"imdb\_data\",\n \"rows\": 49582,\n \"fields\": [\n {\n \"column\": \"review\",\n \"properties\": {\n \"dtype\": \"string\",\n \"num\_unique\_values\": 49582,\n \"samples\": [\n \"\\\"Soul Plane\\\" is a horrible attempt at comedy that only should appeal people with thick skulls, bloodshot eyes and furry pawns. <br /><br />The plot is not only incoherent but also non-existent, acting is mostly sub sub-par with a gang of highly moronic and dreadful characters thrown in for bad measure, jokes are often spotted miles ahead and almost never even a bit amusing. This movie lacks any structure and is full of racial stereotypes that must have seemed old even in the fifties, the only thing it really has going for it is some pretty ladies, but really, if you want that you can rent something from the \\\"Adult\\\" section. OK?<br /><br />I can hardly see anything here to recommend since you'll probably have a lot a better and productive time chasing rats with a sledgehammer or inventing waterproof teabags or whatever.<br /><br />2/10\",\n \"Guest from the Future tells a fascinating story of time travel, friendship, battle of good and evil -- all with a small budget, child actors, and few special effects. Something for Spielberg and Lucas to learn from. ;) A sixth-grader Kolya \\\"Nick\\\" Gerasimov finds a time machine in the basement of a decrepit building and travels 100 years into the future. He discovers a near-perfect, utopian society where robots play guitars and write poetry, everyone is kind to each other and people enjoy everything technology has to offer. Alice is the daughter of a prominent scientist who invented a device called Mielophone that allows to read minds of humans and animals. The device can be put to both good and bad use, depending on whose hands it falls into. When two evil space pirates from Saturn who want to rule the universe attempt to steal Mielophone, it falls into the hands of 20th century school boy Nick. With the pirates hot on his tracks, he travels back to his time, followed by the pirates, and Alice. Chaos, confusion and funny situations follow as the luckless pirates try to blend in with the earthlings. Alice enrolls in the same school Nick goes to and demonstrates superhuman abilities in PE class. The catch is, Alice doesn't know what Nick looks like, while the pirates do. Also, the pirates are able to change their appearance and turn literally into anyone. (Hmm, I wonder if this is where James Cameron got the idea for Terminator...) Who gets to Nick -- and Mielophone -- first? Excellent plot, non-stop adventures, and great soundtrack. I wish Hollywood made kid movies like this one...\",\n \"\\\"National Treasure\\\" (2004) is a thoroughly misguided hodge-podge of plot entanglements that borrow from nearly every cloak and dagger government conspiracy clich\\u00e9 that has ever been written. The film stars Nicholas Cage as Benjamin Franklin Gates (how precious is that, I ask you?); a seemingly normal fellow who, for no other reason than being of a lineage of like-minded misguided fortune hunters, decides to steal a 'national treasure' that has been hidden by the United States founding fathers. After a bit of subtext and background that plays laughably (unintentionally) like Indiana Jones meets The Patriot, the film degenerates into one misguided whimsy after another \\u0096 attempting to create a 'Stanley Goodspeed' regurgitation of Nicholas Cage and launch the whole convoluted mess forward with a series of high octane, but disconnected misadventures.<br /><br />The relevancy and logic to having George Washington and his motley crew of patriots burying a king's ransom someplace on native soil, and then, going through the meticulous plan of leaving clues scattered throughout U.S. currency art work, is something that director Jon Turteltaub never quite gets around to explaining. Couldn't Washington found better usage for such wealth during the start up of the country? Hence, we are left with a mystery built on top of an enigma that is already on shaky ground by the time Ben appoints himself the new custodian of this untold wealth. Ben's intentions are noble \\u0096 if confusing. He's set on protecting the treasure. For who and when?\\u0085your guess is as good as mine.<br /><br />But there are a few problems with Ben's crusade. First up, his friend, Ian Holmes (Sean Bean) decides that he can't wait for Ben to make up his mind about stealing the Declaration of Independence from the National Archives (oh, yeah \\u0096 brilliant idea!). Presumably, the back of that famous document holds the secret answer to the ultimate fortune. So Ian tries to kill Ben. The assassination attempt is, of course, unsuccessful, if overly melodramatic. It also affords Ben the opportunity to pick up, and pick on, the very sultry curator of the archives, Abigail Chase (Diane Kruger). She thinks Ben is clearly a nut \\u0096 at least at the beginning. But true to action/romance form, Abby's resolve melts quicker than you can say, \\\"is that the Hope Diamond?\\\" The film moves into full X-File-ish mode, as the FBI, mistakenly believing that Ben is behind the theft, retaliate in various benign ways that lead to a multi-layering of action sequences reminiscent of Mission Impossible meets The Fugitive. Honestly, don't those guys ever get 'intelligence' information that is correct? In the final analysis, \\\"National Treasure\\\" isn't great film making, so much as it's a patchwork rehash of tired old bits from other movies, woven together from scraps, the likes of which would make IL' Betsy Ross blush.<br /><br />The Buena Vista DVD delivers a far more generous treatment than this film is deserving of. The anamorphic widescreen picture exhibits a very smooth and finely detailed image with very rich colors, natural flesh tones, solid blacks and clean whites. The stylized image is also free of blemishes and digital enhancements. The audio is 5.1 and delivers a nice sonic boom to your side and rear speakers with intensity and realism. Extras include a host of promotional junket material that is rather deep and over the top in its explanation of how and why this film was made. If only, as an audience, we had had more clarification as to why Ben and co. were chasing after an illusive treasure, this might have been one good flick. Extras conclude with the theatrical trailer, audio commentary and deleted scenes. Not for the faint-hearted \\u0096 just the thick-headed.\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n },\n {\n \"column\": \"sentiment\",\n \"properties\": {\n \"dtype\": \"number\",\n \"std\": 0,\n \"min\": 0,\n \"max\": 1,\n \"num\_unique\_values\": 2,\n \"samples\": [\n 0,\n 1\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n },\n {\n \"column\": \"Processed\_review\",\n \"properties\": {\n \"dtype\": \"string\",\n \"num\_unique\_values\": 49577,\n \"samples\": [\n \"i just sat through a very enjoyable fast paced mins of roll roll is about a country boy mat toby malone who has dreams of becoming a sports star mat travels to the city and is to be picked up by his cousin george damien robertson well that was the plan anyway george is involved with a gangster tiny john batchelor and is making a delivery for him needless to say mat gets dragged into george s world i thought it was great how mat teaches george some morals and respect while george teaches mat how to relax and enjoy life a little toby and damien were well cast together and did an outstanding job every character in the movie complimented each other very well the two cops were great david ngoombujarra brought some great comic relief to the movie tiny played a likable gangster that reminded me of one of my favourite characters pando from two hands one of the other things that i liked about roll was that it showcased the cities that i grew up and lived in for years perth and fremantle it was good to see sights and landmarks that i grew up with especially the old ferris wheel this rocks n rolls\",\n \"twenty years ago the five years old boy michael hawthorne witnessed his father killing his mother with an axe in an empty road and committing suicide later on the present days michael gordon currie invites his girlfriend peg stacy grant and his best friends chris myc agnew jennifer emmanuelle vaugier lisa ann kelly benson ned brendon beiser mitch maldive phillip rhys and trish rachel hayward to spend the halloween in the country with his grandparents in their farm he asks his friends to wear costumes that would represent their greatest innermost fear and together with his indian friend crow byron chief moon they would perform an ancient indian celebration using the carved wooden dummy morty jon fedele that would eliminate their fears forever the greatest fear of michael is to become a serial killer like his father but something goes wrong and morty turns into his father killing his friends the fear resurrection is a disappointing and pointless slash movie that uses the interesting concept of eliminating the greatest innermost fear of each friend before it grows but in a messy screenplay full of clich s there are some exaggerated performances like for example ms betsy palmer others very weak but in general the acting is good unfortunately there is no explanation why the dummy is brought to live further in spite of being surrounded by close friends the group does not feel pain or sorrow when each one of them dies the low pace along more than fifty minutes could have been used to built a better dramatic situation in the very end michael shows a charm that his father was interested that i have not noticed along the story i do not know whether the previous reference was edited in the dvd released in brazil with minutes running time the special effects are very reasonable for a b movie my vote is four title brazil fear uma noite de halloween fear one night of halloween\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n }\n ]\n}"

}

},

"metadata": {},

"execution\_count": 54

}

]

},

{

"cell\_type": "code",

"source": [

"lemmatizer = WordNetLemmatizer()\n",

"\n",

"def lemmatize\_review(Procesd\_Colmn):\n",

" Procesd\_Colmn = Procesd\_Colmn.split()\n",

" Procesd\_Colmn = [lemmatizer.lemmatize(word) for word in Procesd\_Colmn]\n",

" Procesd\_Colmn = ' '.join(Procesd\_Colmn)\n",

" return Procesd\_Colmn\n",

"\n",

"imdb\_data['Lemmatized\_review'] = imdb\_data['Processed\_review'].apply(lemmatize\_review)\n",

"\n",

"imdb\_data.head()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 242

},

"id": "pKeuBAmzC0OE",

"outputId": "310554ad-d707-434c-e9fa-a95faf97d738"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"[nltk\_data] Downloading package wordnet to /root/nltk\_data...\n",

"[nltk\_data] Package wordnet is already up-to-date!\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" review sentiment \\\n",

"0 One of the other reviewers has mentioned that ... 1 \n",

"1 A wonderful little production. <br /><br />The... 1 \n",

"2 I thought this was a wonderful way to spend ti... 1 \n",

"3 Basically there's a family where a little boy ... 0 \n",

"4 Petter Mattei's \"Love in the Time of Money\" is... 1 \n",

"\n",

" Processed\_review \\\n",

"0 one of the other reviewers has mentioned that ... \n",

"1 a wonderful little production the filming tech... \n",

"2 i thought this was a wonderful way to spend ti... \n",

"3 basically there s a family where a little boy ... \n",

"4 petter mattei s love in the time of money is a... \n",

"\n",

" Lemmatized\_review \n",

"0 one of the other reviewer ha mentioned that af... \n",

"1 a wonderful little production the filming tech... \n",

"2 i thought this wa a wonderful way to spend tim... \n",

"3 basically there s a family where a little boy ... \n",

"4 petter mattei s love in the time of money is a... "

],

"text/html": [

"\n",

" <div id=\"df-2c89ce27-4bca-4311-a59a-ab393efbab76\" class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>review</th>\n",

" <th>sentiment</th>\n",

" <th>Processed\_review</th>\n",

" <th>Lemmatized\_review</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>One of the other reviewers has mentioned that ...</td>\n",

" <td>1</td>\n",

" <td>one of the other reviewers has mentioned that ...</td>\n",

" <td>one of the other reviewer ha mentioned that af...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>A wonderful little production. &lt;br /&gt;&lt;br /&gt;The...</td>\n",

" <td>1</td>\n",

" <td>a wonderful little production the filming tech...</td>\n",

" <td>a wonderful little production the filming tech...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>I thought this was a wonderful way to spend ti...</td>\n",

" <td>1</td>\n",

" <td>i thought this was a wonderful way to spend ti...</td>\n",

" <td>i thought this wa a wonderful way to spend tim...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>Basically there's a family where a little boy ...</td>\n",

" <td>0</td>\n",

" <td>basically there s a family where a little boy ...</td>\n",

" <td>basically there s a family where a little boy ...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>Petter Mattei's \"Love in the Time of Money\" is...</td>\n",

" <td>1</td>\n",

" <td>petter mattei s love in the time of money is a...</td>\n",

" <td>petter mattei s love in the time of money is a...</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>\n",

" <div class=\"colab-df-buttons\">\n",

"\n",

" <div class=\"colab-df-container\">\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-2c89ce27-4bca-4311-a59a-ab393efbab76')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

"\n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\" viewBox=\"0 -960 960 960\">\n",

" <path d=\"M120-120v-720h720v720H120Zm60-500h600v-160H180v160Zm220 220h160v-160H400v160Zm0 220h160v-160H400v160ZM180-400h160v-160H180v160Zm440 0h160v-160H620v160ZM180-180h160v-160H180v160Zm440 0h160v-160H620v160Z\"/>\n",

" </svg>\n",

" </button>\n",

"\n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" .colab-df-buttons div {\n",

" margin-bottom: 4px;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-2c89ce27-4bca-4311-a59a-ab393efbab76 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-2c89ce27-4bca-4311-a59a-ab393efbab76');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

"\n",

"\n",

"<div id=\"df-4c2184d5-7bf1-4d39-9462-0a1923301296\">\n",

" <button class=\"colab-df-quickchart\" onclick=\"quickchart('df-4c2184d5-7bf1-4d39-9462-0a1923301296')\"\n",

" title=\"Suggest charts\"\n",

" style=\"display:none;\">\n",

"\n",

"<svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <g>\n",

" <path d=\"M19 3H5c-1.1 0-2 .9-2 2v14c0 1.1.9 2 2 2h14c1.1 0 2-.9 2-2V5c0-1.1-.9-2-2-2zM9 17H7v-7h2v7zm4 0h-2V7h2v10zm4 0h-2v-4h2v4z\"/>\n",

" </g>\n",

"</svg>\n",

" </button>\n",

"\n",

"<style>\n",

" .colab-df-quickchart {\n",

" --bg-color: #E8F0FE;\n",

" --fill-color: #1967D2;\n",

" --hover-bg-color: #E2EBFA;\n",

" --hover-fill-color: #174EA6;\n",

" --disabled-fill-color: #AAA;\n",

" --disabled-bg-color: #DDD;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-quickchart {\n",

" --bg-color: #3B4455;\n",

" --fill-color: #D2E3FC;\n",

" --hover-bg-color: #434B5C;\n",

" --hover-fill-color: #FFFFFF;\n",

" --disabled-bg-color: #3B4455;\n",

" --disabled-fill-color: #666;\n",

" }\n",

"\n",

" .colab-df-quickchart {\n",

" background-color: var(--bg-color);\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: var(--fill-color);\n",

" height: 32px;\n",

" padding: 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-quickchart:hover {\n",

" background-color: var(--hover-bg-color);\n",

" box-shadow: 0 1px 2px rgba(60, 64, 67, 0.3), 0 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: var(--button-hover-fill-color);\n",

" }\n",

"\n",

" .colab-df-quickchart-complete:disabled,\n",

" .colab-df-quickchart-complete:disabled:hover {\n",

" background-color: var(--disabled-bg-color);\n",

" fill: var(--disabled-fill-color);\n",

" box-shadow: none;\n",

" }\n",

"\n",

" .colab-df-spinner {\n",

" border: 2px solid var(--fill-color);\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" animation:\n",

" spin 1s steps(1) infinite;\n",

" }\n",

"\n",

" @keyframes spin {\n",

" 0% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" border-left-color: var(--fill-color);\n",

" }\n",

" 20% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 30% {\n",

" border-color: transparent;\n",

" border-left-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 40% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-top-color: var(--fill-color);\n",

" }\n",

" 60% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" }\n",

" 80% {\n",

" border-color: transparent;\n",

" border-right-color: var(--fill-color);\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" 90% {\n",

" border-color: transparent;\n",

" border-bottom-color: var(--fill-color);\n",

" }\n",

" }\n",

"</style>\n",

"\n",

" <script>\n",

" async function quickchart(key) {\n",

" const quickchartButtonEl =\n",

" document.querySelector('#' + key + ' button');\n",

" quickchartButtonEl.disabled = true; // To prevent multiple clicks.\n",

" quickchartButtonEl.classList.add('colab-df-spinner');\n",

" try {\n",

" const charts = await google.colab.kernel.invokeFunction(\n",

" 'suggestCharts', [key], {});\n",

" } catch (error) {\n",

" console.error('Error during call to suggestCharts:', error);\n",

" }\n",

" quickchartButtonEl.classList.remove('colab-df-spinner');\n",

" quickchartButtonEl.classList.add('colab-df-quickchart-complete');\n",

" }\n",

" (() => {\n",

" let quickchartButtonEl =\n",

" document.querySelector('#df-4c2184d5-7bf1-4d39-9462-0a1923301296 button');\n",

" quickchartButtonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

" })();\n",

" </script>\n",

"</div>\n",

"\n",

" </div>\n",

" </div>\n"

],

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "dataframe",

"variable\_name": "imdb\_data",

"summary": "{\n \"name\": \"imdb\_data\",\n \"rows\": 49582,\n \"fields\": [\n {\n \"column\": \"review\",\n \"properties\": {\n \"dtype\": \"string\",\n \"num\_unique\_values\": 49582,\n \"samples\": [\n \"\\\"Soul Plane\\\" is a horrible attempt at comedy that only should appeal people with thick skulls, bloodshot eyes and furry pawns. <br /><br />The plot is not only incoherent but also non-existent, acting is mostly sub sub-par with a gang of highly moronic and dreadful characters thrown in for bad measure, jokes are often spotted miles ahead and almost never even a bit amusing. This movie lacks any structure and is full of racial stereotypes that must have seemed old even in the fifties, the only thing it really has going for it is some pretty ladies, but really, if you want that you can rent something from the \\\"Adult\\\" section. OK?<br /><br />I can hardly see anything here to recommend since you'll probably have a lot a better and productive time chasing rats with a sledgehammer or inventing waterproof teabags or whatever.<br /><br />2/10\",\n \"Guest from the Future tells a fascinating story of time travel, friendship, battle of good and evil -- all with a small budget, child actors, and few special effects. Something for Spielberg and Lucas to learn from. ;) A sixth-grader Kolya \\\"Nick\\\" Gerasimov finds a time machine in the basement of a decrepit building and travels 100 years into the future. He discovers a near-perfect, utopian society where robots play guitars and write poetry, everyone is kind to each other and people enjoy everything technology has to offer. Alice is the daughter of a prominent scientist who invented a device called Mielophone that allows to read minds of humans and animals. The device can be put to both good and bad use, depending on whose hands it falls into. When two evil space pirates from Saturn who want to rule the universe attempt to steal Mielophone, it falls into the hands of 20th century school boy Nick. With the pirates hot on his tracks, he travels back to his time, followed by the pirates, and Alice. Chaos, confusion and funny situations follow as the luckless pirates try to blend in with the earthlings. Alice enrolls in the same school Nick goes to and demonstrates superhuman abilities in PE class. The catch is, Alice doesn't know what Nick looks like, while the pirates do. Also, the pirates are able to change their appearance and turn literally into anyone. (Hmm, I wonder if this is where James Cameron got the idea for Terminator...) Who gets to Nick -- and Mielophone -- first? Excellent plot, non-stop adventures, and great soundtrack. I wish Hollywood made kid movies like this one...\",\n \"\\\"National Treasure\\\" (2004) is a thoroughly misguided hodge-podge of plot entanglements that borrow from nearly every cloak and dagger government conspiracy clich\\u00e9 that has ever been written. The film stars Nicholas Cage as Benjamin Franklin Gates (how precious is that, I ask you?); a seemingly normal fellow who, for no other reason than being of a lineage of like-minded misguided fortune hunters, decides to steal a 'national treasure' that has been hidden by the United States founding fathers. After a bit of subtext and background that plays laughably (unintentionally) like Indiana Jones meets The Patriot, the film degenerates into one misguided whimsy after another \\u0096 attempting to create a 'Stanley Goodspeed' regurgitation of Nicholas Cage and launch the whole convoluted mess forward with a series of high octane, but disconnected misadventures.<br /><br />The relevancy and logic to having George Washington and his motley crew of patriots burying a king's ransom someplace on native soil, and then, going through the meticulous plan of leaving clues scattered throughout U.S. currency art work, is something that director Jon Turteltaub never quite gets around to explaining. Couldn't Washington found better usage for such wealth during the start up of the country? Hence, we are left with a mystery built on top of an enigma that is already on shaky ground by the time Ben appoints himself the new custodian of this untold wealth. Ben's intentions are noble \\u0096 if confusing. He's set on protecting the treasure. For who and when?\\u0085your guess is as good as mine.<br /><br />But there are a few problems with Ben's crusade. First up, his friend, Ian Holmes (Sean Bean) decides that he can't wait for Ben to make up his mind about stealing the Declaration of Independence from the National Archives (oh, yeah \\u0096 brilliant idea!). Presumably, the back of that famous document holds the secret answer to the ultimate fortune. So Ian tries to kill Ben. The assassination attempt is, of course, unsuccessful, if overly melodramatic. It also affords Ben the opportunity to pick up, and pick on, the very sultry curator of the archives, Abigail Chase (Diane Kruger). She thinks Ben is clearly a nut \\u0096 at least at the beginning. But true to action/romance form, Abby's resolve melts quicker than you can say, \\\"is that the Hope Diamond?\\\" The film moves into full X-File-ish mode, as the FBI, mistakenly believing that Ben is behind the theft, retaliate in various benign ways that lead to a multi-layering of action sequences reminiscent of Mission Impossible meets The Fugitive. Honestly, don't those guys ever get 'intelligence' information that is correct? In the final analysis, \\\"National Treasure\\\" isn't great film making, so much as it's a patchwork rehash of tired old bits from other movies, woven together from scraps, the likes of which would make IL' Betsy Ross blush.<br /><br />The Buena Vista DVD delivers a far more generous treatment than this film is deserving of. The anamorphic widescreen picture exhibits a very smooth and finely detailed image with very rich colors, natural flesh tones, solid blacks and clean whites. The stylized image is also free of blemishes and digital enhancements. The audio is 5.1 and delivers a nice sonic boom to your side and rear speakers with intensity and realism. Extras include a host of promotional junket material that is rather deep and over the top in its explanation of how and why this film was made. If only, as an audience, we had had more clarification as to why Ben and co. were chasing after an illusive treasure, this might have been one good flick. Extras conclude with the theatrical trailer, audio commentary and deleted scenes. Not for the faint-hearted \\u0096 just the thick-headed.\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n },\n {\n \"column\": \"sentiment\",\n \"properties\": {\n \"dtype\": \"number\",\n \"std\": 0,\n \"min\": 0,\n \"max\": 1,\n \"num\_unique\_values\": 2,\n \"samples\": [\n 0,\n 1\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n },\n {\n \"column\": \"Processed\_review\",\n \"properties\": {\n \"dtype\": \"string\",\n \"num\_unique\_values\": 49577,\n \"samples\": [\n \"i just sat through a very enjoyable fast paced mins of roll roll is about a country boy mat toby malone who has dreams of becoming a sports star mat travels to the city and is to be picked up by his cousin george damien robertson well that was the plan anyway george is involved with a gangster tiny john batchelor and is making a delivery for him needless to say mat gets dragged into george s world i thought it was great how mat teaches george some morals and respect while george teaches mat how to relax and enjoy life a little toby and damien were well cast together and did an outstanding job every character in the movie complimented each other very well the two cops were great david ngoombujarra brought some great comic relief to the movie tiny played a likable gangster that reminded me of one of my favourite characters pando from two hands one of the other things that i liked about roll was that it showcased the cities that i grew up and lived in for years perth and fremantle it was good to see sights and landmarks that i grew up with especially the old ferris wheel this rocks n rolls\",\n \"twenty years ago the five years old boy michael hawthorne witnessed his father killing his mother with an axe in an empty road and committing suicide later on the present days michael gordon currie invites his girlfriend peg stacy grant and his best friends chris myc agnew jennifer emmanuelle vaugier lisa ann kelly benson ned brendon beiser mitch maldive phillip rhys and trish rachel hayward to spend the halloween in the country with his grandparents in their farm he asks his friends to wear costumes that would represent their greatest innermost fear and together with his indian friend crow byron chief moon they would perform an ancient indian celebration using the carved wooden dummy morty jon fedele that would eliminate their fears forever the greatest fear of michael is to become a serial killer like his father but something goes wrong and morty turns into his father killing his friends the fear resurrection is a disappointing and pointless slash movie that uses the interesting concept of eliminating the greatest innermost fear of each friend before it grows but in a messy screenplay full of clich s there are some exaggerated performances like for example ms betsy palmer others very weak but in general the acting is good unfortunately there is no explanation why the dummy is brought to live further in spite of being surrounded by close friends the group does not feel pain or sorrow when each one of them dies the low pace along more than fifty minutes could have been used to built a better dramatic situation in the very end michael shows a charm that his father was interested that i have not noticed along the story i do not know whether the previous reference was edited in the dvd released in brazil with minutes running time the special effects are very reasonable for a b movie my vote is four title brazil fear uma noite de halloween fear one night of halloween\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n },\n {\n \"column\": \"Lemmatized\_review\",\n \"properties\": {\n \"dtype\": \"string\",\n \"num\_unique\_values\": 49577,\n \"samples\": [\n \"i just sat through a very enjoyable fast paced min of roll roll is about a country boy mat toby malone who ha dream of becoming a sport star mat travel to the city and is to be picked up by his cousin george damien robertson well that wa the plan anyway george is involved with a gangster tiny john batchelor and is making a delivery for him needle to say mat get dragged into george s world i thought it wa great how mat teach george some moral and respect while george teach mat how to relax and enjoy life a little toby and damien were well cast together and did an outstanding job every character in the movie complimented each other very well the two cop were great david ngoombujarra brought some great comic relief to the movie tiny played a likable gangster that reminded me of one of my favourite character pando from two hand one of the other thing that i liked about roll wa that it showcased the city that i grew up and lived in for year perth and fremantle it wa good to see sight and landmark that i grew up with especially the old ferris wheel this rock n roll\",\n \"twenty year ago the five year old boy michael hawthorne witnessed his father killing his mother with an axe in an empty road and committing suicide later on the present day michael gordon currie invite his girlfriend peg stacy grant and his best friend chris myc agnew jennifer emmanuelle vaugier lisa ann kelly benson ned brendon beiser mitch maldive phillip rhys and trish rachel hayward to spend the halloween in the country with his grandparent in their farm he asks his friend to wear costume that would represent their greatest innermost fear and together with his indian friend crow byron chief moon they would perform an ancient indian celebration using the carved wooden dummy morty jon fedele that would eliminate their fear forever the greatest fear of michael is to become a serial killer like his father but something go wrong and morty turn into his father killing his friend the fear resurrection is a disappointing and pointless slash movie that us the interesting concept of eliminating the greatest innermost fear of each friend before it grows but in a messy screenplay full of clich s there are some exaggerated performance like for example m betsy palmer others very weak but in general the acting is good unfortunately there is no explanation why the dummy is brought to live further in spite of being surrounded by close friend the group doe not feel pain or sorrow when each one of them dy the low pace along more than fifty minute could have been used to built a better dramatic situation in the very end michael show a charm that his father wa interested that i have not noticed along the story i do not know whether the previous reference wa edited in the dvd released in brazil with minute running time the special effect are very reasonable for a b movie my vote is four title brazil fear uma noite de halloween fear one night of halloween\"\n ],\n \"semantic\_type\": \"\",\n \"description\": \"\"\n }\n }\n ]\n}"

}

},

"metadata": {},

"execution\_count": 55

}

]

},

{

"cell\_type": "markdown",

"source": [

"### This code creates and displays word clouds for psitiv and negative IMDB reviews. psitiv reviews are shown with warm-colored words on a light blue background, while negative reviews have cool-colored words on a black background. The reviews are first concatenated, and then word clouds are generated using WordCloud, styled, and visualized using matplotlib to highlight frequently used words in each sentiment category."

],

"metadata": {

"id": "14epEyh7VZPy"

}

},

{

"cell\_type": "code",

"source": [

"# Assuming 'sentiment' column contains 1 for psitiv and 0 for negative\n",

"psitiv\_reviews = imdb\_data[imdb\_data['sentiment'] == 1]['Lemmatized\_review']\n",

"negative\_reviews = imdb\_data[imdb\_data['sentiment'] == 0]['Lemmatized\_review']\n",

"\n",

"# Concatenate psitiv and negative reviews into separate strings\n",

"all\_psitiv\_reviews = \" \".join(psitiv\_reviews)\n",

"all\_negative\_reviews = \" \".join(negative\_reviews)\n",

"\n",

"# Customize word cloud appearance for psitiv and negative reviews\n",

"# psitiv: Light blue background, warm colored words\n",

"wordcloud\_psitiv = WordCloud(width=800, height=400,\n",

" background\_color='lightblue',\n",

" colormap='autumn', # Warm colors\n",

" contour\_color='darkorange',\n",

" contour\_width=2).generate(all\_psitiv\_reviews)\n",

"\n",

"# Negative: Dark background, cool colored words\n",

"wordcloud\_negative = WordCloud(width=800, height=400,\n",

" background\_color='black',\n",

" colormap='cool', # Cool colors\n",

" contour\_color='red',\n",

" contour\_width=2).generate(all\_negative\_reviews)\n",

"\n",

"# Display the word clouds with custom styles\n",

"mt\_plot.figure(figsize=(10, 5)) # Adjusting the figure size as needed for good visualization\n",

"mt\_plot.imshow(wordcloud\_psitiv, interpolation='bilinear') # Display the psitiv word cloud\n",

"mt\_plot.axis(\"off\") # Turnd off the axis\n",

"mt\_plot.title('Word Cloud of psitiv IMDB Reviews', fontsize=20, color='darkorange') # Assigning Title and color for psitiv reviews\n",

"mt\_plot.show() # Display the plot thats created by positive words\n",

"\n",

"mt\_plot.figure(figsize=(10, 5)) # Adjusting the figure size as needed for good visualization\n",

"mt\_plot.imshow(wordcloud\_negative, interpolation='bilinear') # Display the negative word cloud\n",

"mt\_plot.axis(\"off\") # Turn off the axis\n",

"mt\_plot.title('Word Cloud of Negative IMDB Reviews', fontsize=20, color='red') # Assigning Title and color for negative reviews\n",

"mt\_plot.show() # Display the plot thats created by negative words\n"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 887

},

"id": "q1tPc263jNLJ",

"outputId": "325b7613-be38-440a-8a53-1db6d8f01347"

},

"execution\_count": null,

"outputs": [

{

"output\_type": "display\_data",

"data": {

"text/plain": [

"<Figure size 1000x500 with 1 Axes>"

],

"image/png": "\n"

},

"metadata": {}

},

{

"output\_type": "display\_data",

"data": {

"text/plain": [

"<Figure size 1000x500 with 1 Axes>"

],

"image/png": "\n"

},

"metadata": {}

}

]

}

]

}