

## Assignment 2

```
In [ ]: import numpy as np
import pandas as pd
from collections import Counter
import functools
from sklearn.metrics import f1_score
```

```
In [ ]: # Reading the document
def read_documents(doc_file) -> tuple:
    docs = []
    labels = []
    with open(doc_file, encoding='utf-8') as f:
        for line in f:
            words = line.strip().split()
            docs.append(words[3:])
            labels.append(words[1])
    return (docs, labels)
```

```
In [ ]: # Dividing into training data and validation data
(all_docs, all_labels) = read_documents('reviews.txt')

# 80% of the data is used for training and 20% for validation.
split_point = int(0.80*len(all_docs))
train_docs = all_docs[:split_point]
train_labels = all_labels[:split_point]
val_docs = all_docs[split_point:]
val_labels = all_labels[split_point:]
#set(train_labels)
#train_docs[0]
print(type(train_labels))

<class 'list'>
```

Estimating parameters for the Naive Bayes classifier

```

In [ ]: '''
    TODO for each word in each document, calculate the probability of that word
    occurring in all different types of labeled documents. Then multiply that with
    the prior probability of some doc being under some label, and lastly divide
    by the total amount of documents of a the particular label.
    start with alpha = 1, as smoothing.
    '''

    # Counts the frequency of each word per label
    def count_words(documents, labels) -> dict:
        label_counts = {}
        for i in range(len(documents)):
            if labels[i] in label_counts.keys():
                label_counts[labels[i]].update(documents[i])
            else:
                label_counts.update({labels[i]:Counter(documents[i])})
        return label_counts

    # Counts the total frequency of all words per label
    def count_total_words(word_count) -> dict:
        words_per_label = {}
        for label in word_count.keys():
            words_per_label.update({label:0})
            for _ in word_count[label]:
                words_per_label[label] += 1
        return words_per_label

    # Trains a Naive Bayes model with given data
    def train_nb(documents, labels) -> dict:
        # The count of each word for each label
        label_counts = count_words(documents, labels)
        label_probs = {}
        # The total amount for each label
        label_word_freq = count_total_words(label_counts)
        for label in label_counts.keys():
            label_probs.update({label:{}})
            for word in label_counts[label]:
                word_count = label_counts[label][word]
                # Adding the probability (converted to logarithmic scale)
                label_probs[label].update(
                    {word:word_count/label_word_freq[label]}
                )
        return label_probs

    res = count_words(train_docs, train_labels);
    print(res['pos']['you'])
    #print([i for i in res['pos']])
    tes = Counter(train_labels)
    #print(tes)
    #res['neg']['i']
    train = train_nb(train_docs, train_labels)
    train
    #print(count_total_words(res))

```

5808

```
Out[1]: {'neg': {'i': 0.4915269196822595,
  'bought': 0.02018240659017358,
  'this': 0.28267137393350983,
  'album': 0.0215651662253604,
  'because': 0.03480435422182995,
  'loved': 0.0036775522212415417,
  'the': 0.9389232127096204,
  'title': 0.003236245954692557,
  'song': 0.010767872903795234,
  '.': 0.9026772580170639,
  'it': 0.36298911444542514,
  "'s": 0.1403059723448073,
  'such': 0.013327449249779346,
  'a': 0.4370991468078847,
  'great': 0.024242424242424242,
  ',': 0.7747278611356281,
  'how': 0.027684613121506326,
  'bad': 0.020859076198882023,
  'can': 0.04054133568696675,
  'rest': 0.005619299794057076,
  'of': 0.3793762871432774,
  'be': 0.09067372756693146,
  'right': 0.010797293321565166,
  '?': 0.04554280670785525,
  'well': 0.023889379229185055,
  'songs': 0.012297734627831715,
  'are': 0.10205942924389526,
  'just': 0.059458664313033247,
  'filler': 0.00088261253309797,
  'and': 0.4698440717858194,
  "n't": 0.12280082377169756,
  'worth': 0.0088261253309797,
  'money': 0.01971167990585466,
  'paid': 0.003000882612533098,
  'for': 0.17749338040600177,
  'either': 0.0073845248602530155,
  'shameless': 0.00023536334215945866,
  'bubblegum': 5.8840835539864665e-05,
  'or': 0.0696969696969697,
  'oversentimentalized': 2.9420417769932333e-05,
  'depressing': 0.000794351279788173,
  'tripe': 0.0002942041776993233,
  'kenny': 0.000176522506619594,
  'chesney': 8.8261253309797e-05,
  'is': 0.30579582230067665,
  'popular': 0.001235657546337158,
  'artist': 0.0015298617240364814,
  'as': 0.10394233598117093,
  'result': 0.0027066784348337746,
  'he': 0.053957046190055896,
  'in': 0.2231832892027067,
  'cookie': 0.000176522506619594,
  'cutter': 0.00011768167107972933,
  'category': 0.000529567519858782,
  'nashville': 0.000176522506619594,
  'music': 0.017181523977640483,
  'scene': 0.0047955280964989706,
  'gotta': 0.0005001471020888497,
  'pump': 0.000176522506619594,
```

'out': 0.05663430420711974,  
'albums': 0.003795233892321271,  
'so': 0.07249190938511327,  
'record': 0.003089143865842895,  
'company': 0.005795822300676669,  
'keep': 0.006854957340394234,  
'lining': 8.8261253309797e-05,  
'their': 0.035775228008237715,  
'pockets': 0.000264783759929391,  
'while': 0.015710503089143867,  
'suckers': 5.8840835539864665e-05,  
'there': 0.052162400706090024,  
'buying': 0.0069432185937040305,  
'garbage': 0.002736098852603707,  
'to': 0.48976169461606356,  
'perpetuate': 5.8840835539864665e-05,  
'more': 0.046572521329802885,  
'coming': 0.0024124742571344513,  
'that': 0.23097969991173875,  
'town': 0.001323918799646955,  
'll': 0.010561929979405707,  
'get': 0.04601353339217417,  
'down': 0.01197411003236246,  
'off': 0.019505736981465138,  
'my': 0.10420711974110032,  
'soapbox': 2.9420417769932333e-05,  
'now': 0.02206531332744925,  
'but': 0.13233303912915564,  
'country': 0.0024124742571344513,  
'really': 0.02836128273021477,  
'needs': 0.004089438070020594,  
'back': 0.022035892909679316,  
'roots': 0.0005884083553986466,  
'stop': 0.0035598705501618125,  
'pop': 0.002471315092674316,  
'nonsense': 0.0009120329508679023,  
'what': 0.05175051485731098,  
'considered': 0.0010885554574874964,  
'by': 0.050603118564283615,  
'mainstream': 0.0006472491909385113,  
'two': 0.02068255369226243,  
'different': 0.008884966166519565,  
'things': 0.008737864077669903,  
'was': 0.15572227125625185,  
'misled': 5.8840835539864665e-05,  
'thought': 0.012444836716681376,  
'entire': 0.005325095616357752,  
'cd': 0.01971167990585466,  
'contains': 0.001235657546337158,  
'one': 0.07602235951750515,  
'have': 0.12544866137099148,  
'introduced': 0.000617828773168579,  
'many': 0.02218299499852898,  
'ell': 2.9420417769932333e-05,  
'high': 0.0065313327449249775,  
'school': 0.0030597234480729627,  
'students': 0.0008531921153280376,  
'lois': 5.8840835539864665e-05,  
'lowery': 2.9420417769932333e-05,

'depth': 0.001059135039717564,  
'her': 0.0313915857605178,  
'characters': 0.00985583995292733,  
'she': 0.026743159752868492,  
'brilliant': 0.001147396293027361,  
'writer': 0.002177110914974993,  
'capable': 0.000794351279788173,  
'inspiring': 0.000264783759929391,  
'fierce': 8.8261253309797e-05,  
'passion': 0.0005589879376287143,  
'readers': 0.0019123271550456017,  
'they': 0.07619888202412474,  
'encounter': 0.00023536334215945866,  
'shocking': 0.000353045013239188,  
'details': 0.0018829067372756693,  
'utopian': 5.8840835539864665e-05,  
'worlds': 0.00023536334215945866,  
'anxious': 0.00011768167107972933,  
'read': 0.020594292438952635,  
'companion': 0.000176522506619594,  
'novel': 0.004148278905560459,  
'had': 0.04830832597822889,  
'planned': 0.00023536334215945866,  
'share': 0.0008531921153280376,  
'with': 0.1450426596057664,  
'class': 0.0018829067372756693,  
'january': 0.00038246543100912034,  
'although': 0.004766107678729038,  
'series': 0.005972344807296264,  
'written': 0.0065313327449249775,  
'6th': 5.8840835539864665e-05,  
'graders': 2.9420417769932333e-05,  
'older': 0.002824360105913504,  
'book': 0.057487496322447776,  
'simplicity': 0.0002942041776993233,  
'its': 0.020152986172403648,  
'message': 0.00353045013239188,  
'language': 0.003000882612533098,  
'writing': 0.005766401882906737,  
'style': 0.0038246543100912034,  
'will': 0.04307149161518094,  
'inspire': 0.000176522506619594,  
'no': 0.05051485731097381,  
'am': 0.02283024418946749,  
'sadly': 0.001235657546337158,  
'disappointed': 0.008708443659899971,  
'experience': 0.0048249485142689025,  
'camera': 0.039952927331568105,  
'takes': 0.007178581935863489,  
'pictures': 0.012797881729920565,  
'zoom': 0.002206531332744925,  
'lens': 0.008708443659899971,  
'delicate': 0.00020594292438952632,  
'breaks': 0.0005589879376287143,  
'easily': 0.002912621359223301,  
'stops': 0.0006472491909385113,  
'extending': 8.8261253309797e-05,  
'makes': 0.009826419535157399,  
'clicking': 0.000441306266548985,

'noise': 0.0025301559282141805,  
'happened': 0.0040305972344807295,  
'me': 0.05225066195939982,  
'when': 0.05007355104442483,  
'accidentally': 0.00038246543100912034,  
'turned': 0.004324801412180053,  
'on': 0.1383053839364519,  
'pants': 0.000353045013239188,  
'pocket': 0.001059135039717564,  
'rendering': 0.0002942041776993233,  
'unusable': 0.0007355104442483084,  
'brother': 0.001235657546337158,  
'friends': 0.0030303030303030303,  
'experienced': 0.0012062371285672256,  
'similar': 0.0027066784348337746,  
'problem': 0.015151515151515152,  
'ex-s500': 5.8840835539864665e-05,  
'after': 0.03486319505736982,  
'dropped': 0.001676963812886143,  
'extended': 0.0010002942041776994,  
'would': 0.06081200353045013,  
'recommend': 0.008178876140041189,  
'looking': 0.008031774051191527,  
'into': 0.02080023536334216,  
'another': 0.01712268314210062,  
'not': 0.15142689026184172,  
'first': 0.02815533980582524,  
'd': 0.009120329508679022,  
'like': 0.06484260076493087,  
'tell': 0.006148867313915857,  
'you': 0.15834068843777582,  
'only': 0.043836422477199176,  
'second': 0.006884377758164166,  
'star': 0.006678434833774639,  
'wars': 0.000617828773168579,  
've': 0.02041776993233304,  
'ever': 0.013680494263018535,  
'': 0.03398058252427184,  
'jedi': 0.0002942041776993233,  
'search': 0.0019711679905854663,  
'being': 0.01362165342747867,  
'actually': 0.012238893792291851,  
'liked': 0.0029420417769932335,  
'drew': 0.00014710208884966165,  
'upon': 0.00264783759929391,  
'lore': 5.8840835539864665e-05,  
'decent': 0.003795233892321271,  
'however': 0.015180935569285084,  
'made': 0.01676963812886143,  
'groan': 2.9420417769932333e-05,  
'places': 0.0011179758752574286,  
'felt': 0.004089438070020594,  
'almost': 0.009532215357458075,  
'compelled': 0.00023536334215945866,  
'create': 0.002559576345984113,  
'an': 0.05566343042071197,  
'account': 0.0013533392174168873,  
'here': 0.014916151809355693,  
'little': 0.01500441306266549,

'rant': 0.000176522506619594,  
'about': 0.054751397469844074,  
'(': 0.08505442777287438,  
'warned': 0.0006766696087084436,  
'may': 0.011032656663724626,  
'contain': 0.0008531921153280376,  
'few': 0.01462194763165637,  
'spoilers': 0.000176522506619594,  
')': 0.08887908208296558,  
'thusly': 2.9420417769932333e-05,  
'::': 0.029832303618711385,  
'much': 0.03153868784936746,  
'wasted': 0.0023536334215945864,  
'overly': 0.000706090026478376,  
'long': 0.010709032068255369,  
'descriptions': 0.0007355104442483084,  
'reminds': 0.000529567519858782,  
'britain': 5.8840835539864665e-05,  
'sketch': 2.9420417769932333e-05,  
'where': 0.01362165342747867,  
'woman': 0.003236245954692557,  
'author': 0.010032362459546926,  
'dictating': 2.9420417769932333e-05,  
'new': 0.022359517505148575,  
'starts': 0.0018240659017358047,  
'reading': 0.009679317446307737,  
'bible': 0.0006472491909385113,  
'make': 0.021418064136510737,  
'up': 0.04136510738452486,  
'space': 0.0015592821418064136,  
'!': 0.08337746395998823,  
'do': 0.07896440129449839,  
'mind': 0.004148278905560459,  
'story': 0.013415710503089144,  
'20-odd': 2.9420417769932333e-05,  
'pages': 0.004942630185348632,  
'shorter': 0.000617828773168579,  
'if': 0.07466902030008826,  
'comes': 0.006295969402765519,  
'point': 0.007914092380111797,  
'some': 0.0411297440423654,  
'bizarre': 0.0006472491909385113,  
'reason': 0.007002059429243895,  
'seems': 0.010032362459546926,  
'taken': 0.0040011768167107975,  
'every': 0.013092085907619889,  
'opportunity': 0.001059135039717564,  
'ground': 0.000794351279788173,  
'constantly': 0.001676963812886143,  
'relating': 0.00011768167107972933,  
'films': 0.003618711385701677,  
'example': 0.004972050603118564,  
'wedge': 5.8840835539864665e-05,  
'scientist': 0.0004707266843189173,  
'go': 0.01588702559576346,  
'forest-resort': 2.9420417769932333e-05,  
'planet': 0.000441306266548985,  
'meet': 0.00088261253309797,  
'alien': 0.00041188584877905265,

'ambassador': 2.9420417769932333e-05,  
'says': 0.005354516034127684,  
'something': 0.014474845542806708,  
'along': 0.0036775522212415417,  
'lines': 0.002177110914974993,  
'cantina': 2.9420417769932333e-05,  
'mos': 2.9420417769932333e-05,  
'eisley': 2.9420417769932333e-05,  
'luke': 0.0002942041776993233,  
'met': 0.0010297146219476316,  
'han': 5.8840835539864665e-05,  
'chewbacca': 2.9420417769932333e-05,  
'did': 0.040129449838187704,  
'know': 0.020476610767872903,  
'brush': 0.0016181229773462784,  
'history': 0.004383642247719918,  
'there...': 2.9420417769932333e-05,  
'gahhhhhhh': 2.9420417769932333e-05,  
'why': 0.015681082671373935,  
'need': 0.012032950867902324,  
'\*bangs': 2.9420417769932333e-05,  
'head': 0.0047955280964989706,  
'desk\*': 2.9420417769932333e-05,  
'does': 0.041070903206825535,  
'throughout': 0.0024124742571344513,  
'books': 0.008855545748749633,  
'far': 0.01032656663724625,  
'stick': 0.002824360105913504,  
'your': 0.039923506913798174,  
'own': 0.010444248308325979,  
'neck': 0.0007355104442483084,  
'come': 0.009944101206237128,  
'without': 0.011238599588114151,  
'proving': 0.0002942041776993233,  
'true': 0.004942630185348632,  
'telling': 0.0021476904972050604,  
'from': 0.06437187408061194,  
'certain': 0.0020888496616651956,  
'then': 0.030273609885260372,  
'admiral': 8.8261253309797e-05,  
'daala': 2.9420417769932333e-05,  
'military': 0.0007355104442483084,  
'skills': 0.0008237716975581053,  
'quick': 0.0015298617240364814,  
'thinking...': 2.9420417769932333e-05,  
'doing': 0.005942924389526331,  
'reviewing': 0.000441306266548985,  
'recordings': 0.0010002942041776994,  
'speeches': 0.00020594292438952632,  
'given': 0.004736687260959106,  
'tarkin': 2.9420417769932333e-05,  
'over': 0.02394822006472492,  
'tactics': 0.00011768167107972933,  
'following': 0.001500441306266549,  
'them': 0.03180347160929685,  
'letter': 0.0005884083553986466,  
'losing': 0.000617828773168579,  
'end': 0.010444248308325979,  
'1': 0.00897322741982936,



'destroyer': 5.8840835539864665e-05,  
'should': 0.019829361576934393,  
'live': 0.005854663136216534,  
'reputation': 0.0005884083553986466,  
'at': 0.06522506619593998,  
'least': 0.008061194468961459,  
'successful': 0.0008237716975581053,  
'campaigns': 5.8840835539864665e-05,  
'before': 0.015445719329214475,  
'inevitable': 0.000264783759929391,  
'happens': 0.00176522506619594,  
'i.e.': 0.001059135039717564,  
'rebellion': 5.8840835539864665e-05,  
'win': 0.0008237716975581053,  
'goes': 0.004883789349808767,  
'waste': 0.009473374521918211,  
'absolutely': 0.003883495145631068,  
'idea': 0.005472197705207414,  
'having': 0.00744336569579288,  
'academy': 0.00032362459546925567,  
'trilogy': 0.0005589879376287143,  
'could': 0.03359811709326273,  
'been': 0.030979699911738746,  
'done': 0.0069137981759340985,  
'-': 0.059576345984112974,  
'original': 0.009414533686378346,  
'training': 0.0011179758752574286,  
'has': 0.049749926448955574,  
'two-by-two': 2.9420417769932333e-05,  
'forest': 0.00038246543100912034,  
'holding': 0.000706090026478376,  
'hands': 0.0018829067372756693,  
'\*sense\*': 2.9420417769932333e-05,  
'things..': 5.8840835539864665e-05,  
'meant': 0.0013827596351868197,  
'sci-fi': 0.0005001471020888497,  
'better': 0.02724330685495734,  
'incensed': 8.8261253309797e-05,  
'find': 0.016799058546631364,  
'boring': 0.006325389820535452,  
'part': 0.007531626949102677,  
'motivation': 0.00020594292438952632,  
'starting': 0.001235657546337158,  
'place': 0.004707266843189173,  
'see': 0.021035598705501618,  
'grown': 0.00038246543100912034,  
'all': 0.07149161518093557,  
'started': 0.005383936451897617,  
'very': 0.046954986760812,  
'promisingly': 2.9420417769932333e-05,  
'learnt': 5.8840835539864665e-05,  
'dark': 0.00353045013239188,  
'side': 0.0029420417769932335,  
'through': 0.01350397175639894,  
'love': 0.013092085907619889,  
'his': 0.04410120623712857,  
'powerful': 0.0013533392174168873,  
'than': 0.03648131803471609,  
'cool': 0.002824360105913504,

'got': 0.019123271550456015,  
'stage': 0.001412180052956752,  
'dread': 0.00011768167107972933,  
'parts': 0.002383053839364519,  
'dull': 0.002206531332744925,  
'other': 0.03324507208002354,  
'trainee': 5.8840835539864665e-05,  
'needed': 0.004207119741100324,  
'serious': 0.0030597234480729627,  
'character': 0.007766990291262136,  
'development': 0.0012650779641070903,  
'were': 0.03233303912915563,  
'dimensional': 0.00032362459546925567,  
'entering': 0.0004707266843189173,  
'surveying': 8.8261253309797e-05,  
'noting': 0.000176522506619594,  
'progressing': 2.9420417769932333e-05,  
'good': 0.04280670785525154,  
'enough': 0.010797293321565166,  
'reviewers': 0.003089143865842895,  
'said': 0.009944101206237128,  
'rogue': 5.8840835539864665e-05,  
'somehow': 0.0015298617240364814,  
'gains': 0.000176522506619594,  
'knowledge': 0.0012650779641070903,  
'building': 0.0009414533686378346,  
'super-lightsabre': 2.9420417769932333e-05,  
'btw': 0.0004707266843189173,  
'we': 0.03086201824065902,  
'discovered': 0.001500441306266549,  
'burnt': 0.00011768167107972933,  
'death': 0.0024418946749043837,  
'quarters..': 2.9420417769932333e-05,  
'meditates': 2.9420417769932333e-05,  
'\*grr\*': 2.9420417769932333e-05,  
'okay': 0.0019711679905854663,  
'strength': 0.000706090026478376,  
'next': 0.006796116504854369,  
'car': 0.0015298617240364814,  
'accident': 0.000264783759929391,  
'look': 0.01085613415710503,  
'even': 0.03968814357163872,  
'finishing': 0.000176522506619594,  
'write': 0.005119152691968226,  
'personal': 0.0025007355104442485,  
'connections': 0.00011768167107972933,  
'blackwater': 5.8840835539864665e-05,  
'employees': 0.00014710208884966165,  
'close': 0.0040011768167107975,  
'sided': 2.9420417769932333e-05,  
'based': 0.0033245072080023536,  
'facts': 0.001235657546337158,  
'which': 0.03589290967931744,  
'most': 0.021888790820829656,  
'writers': 0.0012944983818770227,  
'purely': 0.00023536334215945866,  
'fiction': 0.0018240659017358047,  
'misfortunes': 2.9420417769932333e-05,  
'bases': 0.00014710208884966165,

'postive': 2.9420417769932333e-05,  
'stories': 0.002030008826125331,  
'hear': 0.0040011768167107975,  
'everyday': 0.00041188584877905265,  
'certified': 0.000176522506619594,  
'information': 0.005148573109738159,  
'systems': 0.0007355104442483084,  
'security': 0.002677258017063842,  
'professional': 0.002383053839364519,  
'definitively': 8.8261253309797e-05,  
'state': 0.0033833480435422184,  
'cover': 0.0048249485142689025,  
'everything': 0.00706090026478376,  
'covered': 0.001147396293027361,  
'any': 0.02733156810826714,  
'whatsoever': 0.0007649308620182407,  
'glowing': 0.00032362459546925567,  
'reviews': 0.009561635775228007,  
'found': 0.01350397175639894,  
'm': 0.022035892909679316,  
'loss': 0.001059135039717564,  
'these': 0.02524271844660194,  
'recommended': 0.001853486319505737,  
'aggrevating': 2.9420417769932333e-05,  
'subtitle': 0.00023536334215945866,  
'': 0.19529273315681084,  
'integrated': 0.00020594292438952632,  
'approach': 0.001147396293027361,  
'organization': 0.000529567519858782,  
'lacks': 0.00176522506619594,  
'effective': 0.0009414533686378346,  
'integration': 0.00014710208884966165,  
'techniques': 0.000441306266548985,  
'seem': 0.006060606060606061,  
'address': 0.0011179758752574286,  
'tries': 0.0021476904972050604,  
'half-hearted': 0.00014710208884966165,  
'generally': 0.0014710208884966167,  
'harmless': 0.00011768167107972933,  
'except': 0.00353045013239188,  
'offensive': 0.0005589879376287143,  
'siblings': 2.9420417769932333e-05,  
'hate': 0.002677258017063842,  
'each': 0.007002059429243895,  
'parents': 0.001853486319505737,  
'probably': 0.007972933215651663,  
'funny': 0.004648426007649309,  
'ultra-sophisticated': 2.9420417769932333e-05,  
'cynical': 0.00023536334215945866,  
'adults': 0.0008531921153280376,  
'totally': 0.004089438070020594,  
'inappropriate': 0.00041188584877905265,  
'four': 0.003177405119152692,  
'year': 0.009944101206237128,  
'old': 0.012180052956751985,  
'bring': 0.001853486319505737,  
'house': 0.002912621359223301,  
'slippers': 0.00011768167107972933,  
'comfortable': 0.0004707266843189173,

'stand': 0.002383053839364519,  
'sitting': 0.0013827596351868197,  
'though': 0.010532509561635775,  
's9000': 2.9420417769932333e-05,  
'fits': 0.0007649308620182407,  
'perfectly': 0.0014710208884966167,  
'case': 0.008120035304501325,  
'too': 0.025389820535451604,  
'strap': 0.0011768167107972932,  
'belt': 0.0004707266843189173,  
'loop': 0.000353045013239188,  
'apparently': 0.0027655192703736394,  
'supposed': 0.0048249485142689025,  
'function': 0.0019711679905854663,  
'also': 0.023654015887025596,  
'extremely': 0.0036481318034716093,  
'short': 0.005383936451897617,  
'small': 0.007296263606943219,  
'people': 0.021241541629891145,  
'instead': 0.010120623712856722,  
'returning': 0.001588702559576346,  
'using': 0.012533097969991174,  
'generic': 0.0010885554574874964,  
'protectthe': 2.9420417769932333e-05,  
'adapt': 0.00014710208884966165,  
'work': 0.02709620476610768,  
'bigger': 0.000970873786407767,  
'aftermarket': 2.9420417769932333e-05,  
'take': 0.01576934392468373,  
'seamstress': 5.8840835539864665e-05,  
'sewn': 2.9420417769932333e-05,  
'extensive': 0.00023536334215945866,  
'modification': 5.8840835539864665e-05,  
'design': 0.0031479847013827595,  
'odd': 0.0009120329508679023,  
'appears': 0.002177110914974993,  
'designed': 0.0015592821418064136,  
'someone': 0.00794351279788173,  
'who': 0.03542218299499853,  
'never': 0.020123565754633716,  
'used': 0.018240659017358045,  
'cannon': 0.0005884083553986466,  
'died': 0.001676963812886143,  
'day': 0.009473374521918211,  
'daughters': 0.00038246543100912034,  
'birthday': 0.0006472491909385113,  
'ran': 0.001235657546337158,  
'tried': 0.009061488673139158,  
'bunch': 0.0018829067372756693,  
'cameras': 0.004324801412180053,  
'morning': 0.0016475433951162106,  
'6': 0.004560164754339511,  
'mp': 0.0005001471020888497,  
'3.0': 8.8261253309797e-05,  
'took': 0.008002353633421595,  
'clearer': 0.00011768167107972933,  
'9': 0.001676963812886143,  
'months': 0.007884671962341865,  
'sending': 0.0007355104442483084,

'time': 0.037893498087672846,  
'repair': 0.002471315092674316,  
'power': 0.004265960576640188,  
'button': 0.002059429243895263,  
'stopped': 0.0030303030303030303,  
'working': 0.00656075316269491,  
'easter': 5.8840835539864665e-05,  
'holidays': 0.00011768167107972933,  
'functions': 0.0013533392174168873,  
'buttons': 0.0008531921153280376,  
'switched': 0.000794351279788173,  
'pushed': 0.00038246543100912034,  
'delete': 0.00032362459546925567,  
'menu': 0.0010297146219476316,  
'came': 0.007325684024713151,  
'mode': 0.0023536334215945864,  
'selector': 8.8261253309797e-05,  
'three': 0.00756104736687261,  
'modes': 0.00032362459546925567,  
'say': 0.01626949102677258,  
'sony': 0.00617828773168579,  
'support': 0.01235657546337158,  
'online': 0.0034127684613121504,  
'helpful': 0.0012650779641070903,  
'reliable': 0.0006472491909385113,  
'worst': 0.008267137393350986,  
'textbook': 0.00020594292438952632,  
'encountered': 0.00038246543100912034,  
'during': 0.004677846425419241,  
'undergraduate': 0.00011768167107972933,  
'years': 0.014033539276257723,  
'unfocused': 8.8261253309797e-05,  
'sparse': 0.000176522506619594,  
'completely': 0.006884377758164166,  
'lacking': 0.0010885554574874964,  
'problems': 0.008708443659899971,  
'riggs': 2.9420417769932333e-05,  
'touches': 0.00014710208884966165,  
'briefly': 0.0002942041776993233,  
'lots': 0.0024418946749043837,  
'topics': 0.000706090026478376,  
'treats': 0.00014710208884966165,  
'none': 0.003442188879082083,  
'leaving': 0.001323918799646955,  
'reader': 0.003795233892321271,  
'notion': 0.0002942041776993233,  
'process': 0.0028537805236834362,  
'control': 0.0022653721682847896,  
'works': 0.008296557811120918,  
'real': 0.010032362459546926,  
'life': 0.009591056192997941,  
'big': 0.008649602824360106,  
'fan': 0.006354810238305384,  
'seborg': 5.8840835539864665e-05,  
'et': 0.000441306266548985,  
'al': 0.00038246543100912034,  
'choose': 0.0010885554574874964,  
'between': 0.005884083553986467,  
'week': 0.003618711385701677,

'ago': 0.0052368343630479555,  
'today': 0.0032656663724624888,  
'turn': 0.005119152691968226,  
'dead': 0.003295086790232421,  
'changed': 0.0025889967637540453,  
'batteries': 0.005177993527508091,  
'still': 0.01676963812886143,  
'digital': 0.007208002353633422,  
'going': 0.013415710503089144,  
'europe': 0.0006472491909385113,  
'days': 0.007531626949102677,  
'scramble': 8.8261253309797e-05,  
'replacement': 0.0018829067372756693,  
'grrrrr': 5.8840835539864665e-05,  
'bead': 8.8261253309797e-05,  
'fantasies': 0.000176522506619594,  
'ii': 0.000706090026478376,  
'same': 0.020005884083553988,  
'positive': 0.001853486319505737,  
';': 0.03100912032950868,  
'wish': 0.003971756398940865,  
'looked': 0.0036481318034716093,  
'pretty': 0.008502500735510444,  
'motifs': 5.8840835539864665e-05,  
'incorporate': 5.8840835539864665e-05,  
'beading': 8.8261253309797e-05,  
'projects': 0.0006472491909385113,  
'typed': 0.00014710208884966165,  
'directions': 0.001323918799646955,  
'simplistic': 0.00032362459546925567,  
'diagrams': 0.00023536334215945866,  
'glad': 0.0015298617240364814,  
'feel': 0.008237716975581053,  
'discouraged': 8.8261253309797e-05,  
'trying': 0.010385407472786113,  
'wo': 0.005619299794057076,  
'iii': 0.0004707266843189173,  
'art': 0.0024124742571344513,  
'elegance': 0.00011768167107972933,  
'beadweaving': 2.9420417769932333e-05,  
'coraling': 2.9420417769932333e-05,  
'technique': 0.000529567519858782,  
'remain': 0.000706090026478376,  
'favorite': 0.0023536334215945864,  
'presentation': 0.000529567519858782,  
'computer': 0.011120917917034422,  
'enhanced': 0.00011768167107972933,  
'b&w': 5.8840835539864665e-05,  
'images': 0.003442188879082083,  
'cases': 0.001059135039717564,  
'sort': 0.002794939688143572,  
'added': 0.0017946454839658722,  
'limited': 0.001323918799646955,  
'value': 0.0025889967637540453,  
'video': 0.008620182406590174,  
'library': 0.0013533392174168873,  
'legacy': 0.00041188584877905265,  
'father': 0.001500441306266549,  
'aural': 0.00011768167107972933,

'commentary': 0.0008531921153280376,  
'brannagh': 8.8261253309797e-05,  
'repetitive': 0.0012062371285672256,  
'episode': 0.0009120329508679023,  
'basically': 0.0031479847013827595,  
'repeated': 0.0013533392174168873,  
'6+': 2.9420417769932333e-05,  
'octagenarian': 2.9420417769932333e-05,  
'survivors': 0.000176522506619594,  
'exact': 0.0013533392174168873,  
'comments': 0.0011179758752574286,  
'reflections': 0.000264783759929391,  
'suppose': 0.001500441306266549,  
'receiving': 0.00032362459546925567,  
'refund': 0.00264783759929391,  
'sent': 0.004060017652250662,  
'heard': 0.004736687260959106,  
'anything': 0.011062077081494558,  
'ya': 0.0004707266843189173,  
'ashley': 0.00011768167107972933,  
'judd': 0.00014710208884966165,  
'normal': 0.002177110914974993,  
'locusts': 5.8840835539864665e-05,  
'husband': 0.00176522506619594,  
'played': 0.0032068255369226244,  
'jim': 0.000706090026478376,  
'caviezel': 2.9420417769932333e-05,  
'christ': 0.0007355104442483084,  
'frequency': 0.00023536334215945866,  
'gets': 0.004236540158870256,  
'arrested': 0.00011768167107972933,  
'put': 0.011150338334804354,  
'guilty': 0.00032362459546925567,  
'bench': 0.00014710208884966165,  
'slaughter': 5.8840835539864665e-05,  
'el': 0.000353045013239188,  
'salvador': 8.8261253309797e-05,  
'hires': 5.8840835539864665e-05,  
'always': 0.007002059429243895,  
'superb': 0.0004707266843189173,  
'morgan': 2.9420417769932333e-05,  
'freeman': 8.8261253309797e-05,  
'unleashed': 8.8261253309797e-05,  
'million': 0.0010002942041776994,  
'dollar': 0.0009120329508679023,  
'baby': 0.0025889967637540453,  
'help': 0.00844365989997058,  
'truth': 0.0021476904972050604,  
'imprisonment': 2.9420417769932333e-05,  
'threats': 0.00011768167107972933,  
'opening': 0.001412180052956752,  
'sequence': 0.000617828773168579,  
'pulls': 0.00014710208884966165,  
'thrills': 5.8840835539864665e-05,  
'mixed': 0.0008237716975581053,  
'bag': 0.0021476904972050604,  
'crash': 0.001059135039717564,  
'outcome': 0.000264783759929391,  
'saw': 0.004383642247719918,

'anyway': 0.002471315092674316,  
'thrid': 5.8840835539864665e-05,  
'thriller': 0.0008237716975581053,  
'murderous': 0.00011768167107972933,  
'manipulative': 0.00011768167107972933,  
'double': 0.0010885554574874964,  
'jeopardy': 0.00011768167107972933,  
'twisted': 0.0005884083553986466,  
'kiss': 0.00014710208884966165,  
'girls': 0.0028537805236834362,  
'killer': 0.0010297146219476316,  
'running': 0.003706972639011474,  
'playing': 0.003353927625772286,  
'games': 0.0014416004707266843,  
'noggin': 2.9420417769932333e-05,  
'starring': 0.000441306266548985,  
'bruce': 0.0005884083553986466,  
'davison': 2.9420417769932333e-05,  
'x-men': 0.00014710208884966165,  
'dahmer': 0.00011768167107972933,  
'amanda': 8.8261253309797e-05,  
'peet': 2.9420417769932333e-05,  
'whole': 0.007472786113562812,  
'nine': 0.00032362459546925567,  
'yards': 0.000353045013239188,  
'ten': 0.001412180052956752,  
'tom': 0.0014710208884966167,  
'bower': 2.9420417769932333e-05,  
'die': 0.001500441306266549,  
'hard': 0.01235657546337158,  
'2': 0.014769049720506031,  
'kmfdm': 8.8261253309797e-05,  
'release..': 5.8840835539864665e-05,  
'esch': 2.9420417769932333e-05,  
'schulz': 2.9420417769932333e-05,  
'sascha': 2.9420417769932333e-05,  
'skold': 5.8840835539864665e-05,  
'messin': 2.9420417769932333e-05,  
'around': 0.008414239482200648,  
'pig.': 2.9420417769932333e-05,  
'listen': 0.004766107678729038,  
'once': 0.007208002353633422,  
'boring.': 2.9420417769932333e-05,  
'dirty': 0.0009120329508679023,  
'skurk': 2.9420417769932333e-05,  
'preach': 0.00014710208884966165,  
'/' : 0.02506619593998235,  
'pervert': 8.8261253309797e-05,  
'listening': 0.003177405119152692,  
'embarrassed': 8.8261253309797e-05,  
'risen': 5.8840835539864665e-05,  
'lyrically': 0.000264783759929391,  
'horrible': 0.005354516034127684,  
'wrong': 0.007737569873492204,  
'defense': 0.00038246543100912034,  
'badass': 2.9420417769932333e-05,  
'voice': 0.0052368343630479555,  
'solo': 0.0009120329508679023,  
'him': 0.011797587525742866,



'symbols': 8.8261253309797e-05,  
'adios': 8.8261253309797e-05,  
'mdfmk': 2.9420417769932333e-05,  
'attak': 2.9420417769932333e-05,  
'bit': 0.007119741100323625,  
'bland': 0.0007649308620182407,  
'times': 0.011503383348043543,  
'think': 0.01879964695498676,  
'urban': 0.00032362459546925567,  
'monkey': 0.000176522506619594,  
'warfare': 5.8840835539864665e-05,  
'cool..': 2.9420417769932333e-05,  
'whatever': 0.002118270079435128,  
'guy': 0.005119152691968226,  
'buty': 2.9420417769932333e-05,  
'awesome': 0.0011768167107972932,  
'sorry.': 5.8840835539864665e-05,  
'colorvision': 0.00011768167107972933,  
'spyder2': 8.8261253309797e-05,  
'monitor': 0.0013827596351868197,  
'calibrated': 0.00014710208884966165,  
'free': 0.006148867313915857,  
'adobe': 0.0032656663724624888,  
'gamma': 8.8261253309797e-05,  
'program': 0.011444542512503677,  
'installed': 0.0038246543100912034,  
'photoshop': 0.000794351279788173,  
'photos': 0.004707266843189173,  
'printed': 0.0007355104442483084,  
'colors': 0.0008531921153280376,  
'spot': 0.0005884083553986466,  
'express': 0.0007649308620182407,  
'looks': 0.005383936451897617,  
'worse': 0.0052368343630479555,  
'dim': 0.00032362459546925567,  
'color': 0.002794939688143572,  
'temperature': 0.0005589879376287143,  
'warm': 0.001235657546337158,  
'green': 0.0010885554574874964,  
'known': 0.0025889967637540453,  
'ruin': 0.00038246543100912034,  
'already': 0.0065313327449249775,  
'hope': 0.004148278905560459,  
'satisfied': 0.0007355104442483084,  
'product': 0.029185054427772875,  
'try': 0.01085613415710503,  
'calibration': 5.8840835539864665e-05,  
'software': 0.01703442188879082,  
'tools': 0.0007355104442483084,  
'intersesting': 2.9420417769932333e-05,  
'topic': 0.0007649308620182407,  
'studying': 0.00032362459546925567,  
'impending': 0.00020594292438952632,  
'world': 0.007855251544571933,  
'oil': 0.0008237716975581053,  
'shortage': 0.00011768167107972933,  
'run': 0.005472197705207414,  
'tells': 0.0014710208884966167,  
'hubbert': 8.8261253309797e-05,

'analysis': 0.0010002942041776994,  
'talk': 0.00441306266548985,  
'lot': 0.011709326272433068,  
'drilling': 2.9420417769932333e-05,  
'20': 0.003442188879082083,  
'drills': 2.9420417769932333e-05,  
'sometimes': 0.003618711385701677,  
'size': 0.0035010297146219477,  
'fields': 0.0005001471020888497,  
'reexamine': 2.9420417769932333e-05,  
'later': 0.006325389820535452,  
'rate': 0.0018240659017358047,  
'plots': 0.000529567519858782,  
'brought': 0.0014710208884966167,  
'driving': 0.00088261253309797,  
'less': 0.008031774051191527,  
'future': 0.00264783759929391,  
'fossil': 2.9420417769932333e-05,  
'fuels': 2.9420417769932333e-05,  
'discuss': 0.000264783759929391,  
'elctricity': 2.9420417769932333e-05,  
'natural': 0.0014416004707266843,  
'gases': 2.9420417769932333e-05,  
'runs': 0.0014710208884966167,  
'stuff': 0.005177993527508091,  
'man': 0.007149161518093557,  
'gas': 0.0002942041776993233,  
'companys': 2.9420417769932333e-05,  
'stay': 0.0032656663724624888,  
'quiet': 0.000529567519858782,  
'want': 0.01827007943512798,  
'told': 0.005325095616357752,  
'everybody': 0.0007649308620182407,  
'ver': 5.8840835539864665e-05,  
'knowledgable': 0.00011768167107972933,  
'knew': 0.0026184171815239778,  
'talking': 0.0025889967637540453,  
'anybody': 0.0006766696087084436,  
'unless': 0.004913209767578699,  
'smart': 0.0008237716975581053,  
'again': 0.01400411885848779,  
'downfall': 0.00020594292438952632,  
'went': 0.006443071491615181,  
'detail': 0.0017063842306560752,  
'gave': 0.005795822300676669,  
'low': 0.004501323918799647,  
'rating': 0.0025889967637540453,  
'attenion': 2.9420417769932333e-05,  
'span': 0.00014710208884966165,  
'recently': 0.002118270079435128,  
'purchased': 0.008473080317740512,  
'acrobat': 0.0011179758752574286,  
'8.0': 0.00020594292438952632,  
'fine': 0.006766696087084437,  
'vista': 0.004265960576640188,  
'machine': 0.003089143865842895,  
'cites': 8.8261253309797e-05,  
'office': 0.0040011768167107975,  
'2003': 0.001588702559576346,

```
'products': 0.005501618122977346,  
'ms': 0.002383053839364519,  
'2007.': 0.000706090026478376,  
'installation': 0.0013533392174168873,  
'reinstallation': 5.8840835539864665e-05,  
'conversion': 0.00038246543100912034,  
'options': 0.001147396293027361,  
'show': 0.00756104736687261,  
'word': 0.005531038540747279,  
'use': 0.026978523095027948,  
'directly': 0.0014416004707266843,  
'file': 0.0027066784348337746,  
'reinstalls': 5.8840835539864665e-05,  
'itself': 0.004177699323330391,  
'reports': 0.000706090026478376,  
'several': 0.008884966166519565,  
'corrupt': 0.00041188584877905265,  
'ask': 0.0025301559282141805,  
'colleagues': 8.8261253309797e-05,  
'pdf': 0.0010885554574874964,  
'plans': 0.00032362459546925567,  
'compatible': 0.001676963812886143,  
'2007': 0.003295086790232421,  
'3.2': 5.8840835539864665e-05,  
'ghz': 0.00011768167107972933,  
'gb': 0.0002942041776993233,  
'pc': 0.0040305972344807295,  
'clearly': 0.0025889967637540453,  
'ready': 0.0013827596351868197,  
'compatibility': 0.0002942041776993233,  
'wizard': 0.000176522506619594,  
'uninstalling': 0.000529567519858782,  
'conflict': 0.000529567519858782,  
'optional': 0.00014710208884966165,  
'conflicts': 0.0002942041776993233,  
'install': 0.003854074727861136,  
'yet': 0.007119741100323625,  
'didn't': 0.000264783759929391,  
'hour': 0.0030597234480729627,  
'installing': 0.0012062371285672256,  
'fails': 0.0019711679905854663,  
'restores': 5.8840835539864665e-05,  
'xp': 0.004883789349808767,  
'realise': 0.000264783759929391,  
'mine': 0.0025889967637540453,  
'isolated': 0.00014710208884966165,  
'typical': 0.0011179758752574286,  
'microsoft': 0.005383936451897617,  
'buy': 0.023595175051485732,  
'mac': 0.0032656663724624888,  
'phone': 0.002824360105913504,  
'800': 0.00020594292438952632,  
'number': 0.00441306266548985,  
'tech-support': 5.8840835539864665e-05,  
'distance': 0.000529567519858782,  
'call': 0.004736687260959106,  
...},  
'pos': {'anything': 0.006220578345662408,  
'you': 0.16274377942165433,
```

'purchase': 0.005323918403945304,  
'in': 0.238231338264963,  
'the': 0.9387469177314504,  
'left': 0.003978928491369648,  
'behind': 0.0020174848688634837,  
'series': 0.00829410446088321,  
'is': 0.3323806321452589,  
'an': 0.059011432414256895,  
'excellent': 0.012357094821788837,  
'read': 0.018549652544272585,  
'.': 0.8400863035193903,  
'these': 0.024798251513113653,  
'books': 0.007117238287379511,  
'are': 0.10202308899349921,  
'great': 0.05626541134274826,  
'and': 0.5254987670925801,  
'very': 0.057442277516251965,  
'close': 0.004763505940372114,  
'to': 0.44488343420757676,  
'bible': 0.0014850930284689532,  
'i': 0.4150975117686617,  
'have': 0.11589329746693566,  
'entire': 0.003110289172831204,  
'set': 0.01045169244563999,  
'amazon': 0.005660165882089218,  
'a': 0.45348576552342523,  
'shopping': 0.0005323918403945303,  
'site': 0.0015411342748262722,  
'they': 0.05099753418516028,  
'ship': 0.0008966599417171038,  
'fast': 0.005295897780766644,  
'would': 0.040629903609056266,  
'recommend': 0.014262497197937683,  
'any': 0.02381752970186057,  
'christian': 0.0011488455503250391,  
'wanting': 0.0010928043039677203,  
'know': 0.01485093028468953,  
'about': 0.047691100650078457,  
'what': 0.03869648060972876,  
'expect': 0.0026339385787939925,  
'during': 0.005183815288052006,  
'return': 0.0022416498542927594,  
'of': 0.39808899349921545,  
'christ': 0.0004763505940372114,  
'!': 0.07730889934992154,  
'fiction': 0.0011488455503250391,  
'but': 0.10796346110737502,  
'still': 0.02269670477471419,  
'makes': 0.01218897108271688,  
'good': 0.04584173952028693,  
'point': 0.007817753866845998,  
'loved': 0.005071732795337369,  
'movies': 0.006500784577449003,  
',': 0.7714637973548532,  
'cant': 0.0005043712172158709,  
'wiat': 2.8020623178659493e-05,  
'for': 0.1879343196592692,  
'third': 0.0019614436225061645,  
'one': 0.07557162071284465,

'funny': 0.0034465366509751176,  
'not': 0.0929724277067922,  
'suitable': 0.00028020623178659493,  
'children': 2.8020623178659493e-05,  
'this': 0.2631416722707913,  
'shaver': 0.002269670477471419,  
'private': 0.0007565568258238063,  
'area': 0.0019614436225061645,  
'however': 0.012833445415826048,  
'key': 0.0022416498542927594,  
'getting': 0.0073694238959874465,  
'best': 0.029169468728984533,  
'results': 0.0038948666218336695,  
'trim': 0.0003922887245012329,  
'longer': 0.003222371665545842,  
'hairs': 0.0004483299708585519,  
'with': 0.15257229320780094,  
'scissors': 8.406186953597848e-05,  
'or': 0.05640551445864156,  
'largest': 0.00019614436225061645,  
'guard': 0.00036426810132257344,  
'first': 0.029449674960771128,  
'will': 0.042899574086527684,  
'keep': 0.008490248823133827,  
'from': 0.06497982515131137,  
'pulling': 0.0004203093476798924,  
'on': 0.12567249495628782,  
'enable': 0.00014010311589329746,  
'foil': 0.00028020623178659493,  
'part': 0.008378166330419188,  
'work': 0.01944631248598969,  
'be': 0.07559964133602332,  
'able': 0.00835014570724053,  
'do': 0.05374355525666891,  
'its': 0.02216431293431966,  
'job': 0.008154001344989913,  
'if': 0.05864716431293432,  
'too': 0.01810132257341403,  
'long': 0.012020847343644923,  
'only': 0.03519390271239632,  
'problem': 0.00770567137413136,  
'had': 0.036174624523649404,  
'was': 0.10970073974445192,  
'that': 0.1937626092804304,  
'it': 0.32862586863931853,  
'did': 0.02370544720914593,  
'me': 0.04068594485541358,  
'shave': 0.002493835462900695,  
'my': 0.10759919300605246,  
'back': 0.01664425016812374,  
'like': 0.05169804976462677,  
'claimed': 0.00019614436225061645,  
'use': 0.03679107823357992,  
'': 0.03267204662631697,  
'mangroomer': 2.8020623178659493e-05,  
'perfect': 0.007565568258238063,  
'off': 0.01322573414032728,  
'all': 0.06803407307778525,  
'your': 0.039396996189195246,

'hair': 0.00770567137413136,  
'easily': 0.005744227751625196,  
'elongated': 2.8020623178659493e-05,  
'handle': 0.0014010311589329747,  
'product': 0.02090338489127998,  
'as': 0.12435552566689083,  
'well': 0.03578233579914817,  
'therefore': 0.0011768661735036988,  
'say': 0.014094373458865725,  
'two': 0.01880183815288052,  
'products': 0.004034969737726967,  
'coupled': 0.00022416498542927594,  
'together': 0.005772248374803855,  
'seem': 0.004735485317193454,  
'cover': 0.0029141448105805875,  
'bases': 0.00019614436225061645,  
'men': 0.0028581035642232682,  
's': 0.14876148845550324,  
'grooming': 8.406186953597848e-05,  
'body': 0.0033624747814391394,  
'highly': 0.00994732122842412,  
'both': 0.012917507285362026,  
'them': 0.026899798251513115,  
'manscaping': 2.8020623178659493e-05,  
're': 0.011152208025106478,  
'intimidated': 0.00016812373907195696,  
'by': 0.05057722483748039,  
'tolstoy': 8.406186953597848e-05,  
'dostoevsky': 5.6041246357318986e-05,  
'masterpieces': 0.00011208249271463797,  
'consider': 0.0021295673615781214,  
'starting': 0.001933422999327505,  
'instead': 0.004651423447657476,  
'turgenev': 2.8020623178659493e-05,  
'': 0.1761936785474109,  
'fathers': 0.00011208249271463797,  
'sons': 0.00030822685496525445,  
'book': 0.04822349249047299,  
'course': 0.00720130015691549,  
'masterpiece': 0.0012609280430396772,  
'russian': 0.0003922887245012329,  
'literature': 0.0007565568258238063,  
'characters': 0.006276619592019726,  
'themes': 0.0009527011880744228,  
'timeless': 0.0006724949562878278,  
'absolutely': 0.0038388253754763507,  
'relatable': 5.6041246357318986e-05,  
'our': 0.011684599865501008,  
'modern': 0.0018213405066128672,  
'culture': 0.0016251961443622506,  
'story': 0.014178435328401703,  
'family': 0.007313382649630128,  
'relationships': 0.0011208249271463797,  
'romance': 0.0011208249271463797,  
'philosophy': 0.0008125980721811253,  
'recommended': 0.004259134723156243,  
'p.s.': 0.00019614436225061645,  
'when': 0.043431965926922216,  
'done': 0.00773369199731002,

'here': 0.012973548531719345,  
'try': 0.005856310244339834,  
'war': 0.005351939027123964,  
'peace': 0.0012329074198610178,  
'at': 0.06262609280430396,  
'least': 0.00608047522976911,  
'crime': 0.0009527011880744228,  
'punishment': 0.00019614436225061645,  
'martin': 0.0007285362026451469,  
'short': 0.005295897780766644,  
'priceless': 0.0004203093476798924,  
'movie': 0.027264066352835688,  
'-': 0.05598520511096167,  
'enjoyed': 0.004231114099977584,  
'every': 0.014150414705223045,  
'minute': 0.001877381752970186,  
'just': 0.04673839946200403,  
'got': 0.012357094821788837,  
'case': 0.007453485765523425,  
'fabric': 0.00036426810132257344,  
'softener': 2.8020623178659493e-05,  
'last': 0.010087424344317418,  
'week': 0.0024097735933647163,  
'yesterday': 0.00036426810132257344,  
'walking': 0.0007285362026451469,  
'downstairs': 0.00011208249271463797,  
'smell': 0.0006444743331091683,  
'lavender': 0.00022416498542927594,  
'hit': 0.003110289172831204,  
'surprised': 0.002269670477471419,  
'because': 0.025134498991257567,  
'n't': 0.08039116789957408,  
'sprayed': 5.6041246357318986e-05,  
'freshener': 2.8020623178659493e-05,  
'air': 0.0015131136516476126,  
'realized': 0.0007845774490024658,  
'boyfriend': 0.0006444743331091683,  
'doing': 0.004791526563550774,  
'load': 0.0008125980721811253,  
'laundry': 0.00019614436225061645,  
'seventh': 0.00030822685496525445,  
'generation': 0.0012889486662183366,  
'detergent': 8.406186953597848e-05,  
'love': 0.02771239632369424,  
'tandem': 2.8020623178659493e-05,  
'clothes': 0.00025218560860793546,  
'\*real\*': 5.6041246357318986e-05,  
'(': 0.09675521183591124,  
'chemically-based': 2.8020623178659493e-05,  
'smells': 0.00030822685496525445,  
'mainstream': 0.0005323918403945303,  
')': 0.09972539789284914,  
'come': 0.008462228199955167,  
'out': 0.043880295897780766,  
'clean': 0.003026227303295225,  
'soft': 0.0021295673615781214,  
'especially': 0.008714413808563103,  
'sheets': 0.00014010311589329746,  
'blankets': 2.8020623178659493e-05,

'gentle': 0.0006444743331091683,  
'enough': 0.010171486213853395,  
'baby': 0.002661959201972652,  
'made': 0.01221699170589554,  
'natural': 0.0014010311589329747,  
'real': 0.008041918852275275,  
'essential': 0.0014290517821116341,  
'oil': 0.001597175521183591,  
'also': 0.03564223268325487,  
'safe': 0.0007565568258238063,  
'septic': 2.8020623178659493e-05,  
'system': 0.0062485989688410674,  
'so': 0.06128110289172831,  
"'m": 0.018129343196592693,  
'sold': 0.0010367630576104013,  
'sensitive': 0.0015691548980049316,  
'chemical-based': 2.8020623178659493e-05,  
'scents': 0.00011208249271463797,  
'softeners': 2.8020623178659493e-05,  
'ca': 0.010059403721138758,  
'find': 0.016420085182694465,  
'outstanding': 0.001933422999327505,  
'camera': 0.038948666218336696,  
'respect': 0.0007565568258238063,  
"'ve": 0.023957632817753868,  
'owned': 0.002101546738399462,  
'several': 0.007985877605917955,  
'digital': 0.008770455054920422,  
'cameras': 0.0042871553463349026,  
'recent': 0.0015691548980049316,  
'years': 0.01989464245684824,  
'sony': 0.0028861241874019276,  
'&': 0.019474333109168348,  
'nikon': 0.0023537323470073975,  
'n2': 2.8020623178659493e-05,  
'giant': 0.0005323918403945303,  
'3': 0.0073414032728087875,  
'lcd': 0.00165321676754091,  
'touch': 0.0017652992602555481,  
'screen': 0.005576104012553239,  
'features': 0.009078681909885675,  
'lens': 0.019138085631024433,  
'optics': 0.00030822685496525445,  
'overall': 0.006668908316520959,  
'ease': 0.0016251961443622506,  
'top': 0.005351939027123964,  
'previous': 0.003978928491369648,  
'land': 0.0009807218112530822,  
'underwater': 0.0005604124635731899,  
'photography': 0.002549876709258014,  
'performs': 0.0005043712172158709,  
'flawlessly': 0.0005604124635731899,  
'mom': 0.0011208249271463797,  
'frame': 0.0020174848688634837,  
'husband': 0.003026227303295225,  
'easy': 0.019726518717776285,  
'download': 0.0012609280430396772,  
'pictures': 0.012357094821788837,  
'she': 0.021912127325711725,



'can': 0.0529869984308451,  
'show': 0.007677650750952701,  
'her': 0.02925353059852051,  
'grandbabies': 2.8020623178659493e-05,  
'awsome': 0.00016812373907195696,  
'gift': 0.0020735261152208025,  
'anyone': 0.009807218112530823,  
'color': 0.0038948666218336695,  
'believe': 0.0054360008966599415,  
'\$1000.': 8.406186953597848e-05,  
'forget': 0.002157587984756781,  
'zoom': 0.005800268997982515,  
'lenses': 0.003754763505940372,  
'even': 0.027376148845550324,  
'l': 0.0011768661735036988,  
'far': 0.010703878054247927,  
'better': 0.020819323021744003,  
'than': 0.03474557274153777,  
'70-200l': 2.8020623178659493e-05,  
'opinion': 0.003054247926473885,  
'sharp': 0.0021856086079354406,  
'f2': 8.406186953597848e-05,  
'more': 0.052762833445415824,  
'f2.5.': 5.6041246357318986e-05,  
'f': 0.0016812373907195697,  
'/': 0.02547074646940148,  
'5.6': 2.8020623178659493e-05,  
'result': 0.0021856086079354406,  
'8': 0.0023537323470073975,  
'match': 0.0010367630576104013,  
'built': 0.0019614436225061645,  
'excessive': 0.00022416498542927594,  
'heavy': 0.001989464245684824,  
'hood': 0.0017933198834342075,  
'bag': 0.0035305985205110963,  
'unfortunately': 0.0013449899125756557,  
'canon': 0.007957856982739295,  
'decided': 0.00330643353508182,  
'remove': 0.0011768661735036988,  
'hard': 0.008041918852275275,  
'cases': 0.0010367630576104013,  
'combined': 0.0007845774490024658,  
'tamron': 0.00014010311589329746,  
'1.4x': 0.00030822685496525445,  
'teleconverter': 0.00014010311589329746,  
'provide': 0.00165321676754091,  
'optical': 0.0017372786370768885,  
'quality': 0.014766868415153553,  
'par': 0.00028020623178659493,  
'189mm': 2.8020623178659493e-05,  
'focal': 0.0006444743331091683,  
'length': 0.0017372786370768885,  
'yes': 0.004091010984084286,  
'lost': 0.0037827841291190314,  
'convenience': 0.0005323918403945303,  
'70-200': 0.00025218560860793546,  
'am': 0.020539116789957408,  
'really': 0.03286819098856759,  
'happy': 0.006220578345662408,

'decide': 0.0011768661735036988,  
'change': 0.004343196592692221,  
'plan': 0.0021295673615781214,  
'buying': 0.0036426810132257344,  
'70-200is': 2.8020623178659493e-05,  
'135': 0.00011208249271463797,  
'2l+1.4x': 2.8020623178659493e-05,  
'saves': 0.0005604124635731899,  
'money': 0.00717327953373683,  
'let': 0.005379959650302623,  
'get': 0.03667899574086528,  
'wide': 0.0029701860569379063,  
'open': 0.0037827841291190314,  
'portrait': 0.0006724949562878278,  
'200': 0.0008966599417171038,  
'1.8': 0.00016812373907195696,  
'300': 0.0007565568258238063,  
'2.8': 0.0006164537099305089,  
'cost': 0.00277404169468729,  
'thousands': 0.0007285362026451469,  
'dollars': 0.0011768661735036988,  
'trust': 0.0011208249271463797,  
'once': 0.008378166330419188,  
'play': 0.006977135171486214,  
'primes': 0.00011208249271463797,  
'anymore': 0.0010087424344317419,  
'using': 0.01541134274826272,  
'24-70l': 5.6041246357318986e-05,  
'shoot': 0.003110289172831204,  
'small': 0.009134723156242995,  
'kid': 0.00165321676754091,  
'after': 0.023537323470073975,  
'thinking': 0.0023537323470073975,  
'someday': 0.00028020623178659493,  
'reason': 0.0048475678099080925,  
'preventing': 0.00014010311589329746,  
'1.6x': 0.00019614436225061645,  
'crop': 0.0005604124635731899,  
'factor': 0.0007565568258238063,  
'35mm': 0.0009246805648957633,  
'1.4l': 2.8020623178659493e-05,  
'become': 0.004175072853620265,  
'56mm': 2.8020623178659493e-05,  
'which': 0.03510984084286035,  
'order': 0.004034969737726967,  
'free': 0.003922887245012329,  
'shipping': 0.0012048867966823582,  
'though': 0.012020847343644923,  
'says': 0.0033904954046177988,  
'stock': 0.0009527011880744228,  
'shipped': 0.0004763505940372114,  
'mine': 0.002045505492042143,  
'2': 0.011292311140999775,  
'days': 0.0059683927370544725,  
'happened': 0.0015411342748262722,  
'before': 0.014038332212508406,  
'item': 0.00274602107150863,  
'stated': 0.0007285362026451469,  
'soon': 0.0026899798251513113,

'pros': 0.0011208249271463797,  
' ': 0.02807666442501681,  
'super': 0.0016812373907195697,  
'start': 0.006164537099305089,  
'bokeh': 0.0003922887245012329,  
'background': 0.0022976911006500783,  
'blurring': 0.00014010311589329746,  
'extremely': 0.0034465366509751176,  
'picture': 0.0070611970410221925,  
'contrast': 0.0011208249271463797,  
'smaller': 0.0017372786370768885,  
'size': 0.00549204214301726,  
'less': 0.00720130015691549,  
'weight': 0.002045505492042143,  
'black': 0.003922887245012329,  
'compare': 0.0015691548980049316,  
'those': 0.016560188298587762,  
'white': 0.0036146603900470745,  
'70-200s': 2.8020623178659493e-05,  
'build': 0.0014290517821116341,  
'quiet': 0.0010367630576104013,  
'ring': 0.0004203093476798924,  
'usm': 0.0006444743331091683,  
'af': 0.0003362474781439139,  
'ftm': 2.8020623178659493e-05,  
'cons': 0.0011768661735036988,  
'no': 0.031214974221026676,  
'inexpensive': 0.0006724949562878278,  
'effective': 0.0015131136516476126,  
'216mm': 2.8020623178659493e-05,  
'10d': 0.00014010311589329746,  
'20d': 0.0006444743331091683,  
'rebel': 0.001989464245684824,  
'300d': 8.406186953597848e-05,  
'little': 0.018381528805200626,  
'difficult': 0.0033624747814391394,  
'low': 0.00495965030262273,  
'tech': 0.0007845774490024658,  
'brilliant': 0.002269670477471419,  
'music': 0.01995068370320556,  
'seems': 0.008462228199955167,  
'channeled': 2.8020623178659493e-05,  
'zappa': 2.8020623178659493e-05,  
'winchester': 5.6041246357318986e-05,  
'cathedral': 0.00011208249271463797,  
'god': 0.0027180004483299707,  
'knows': 0.0018493611297915266,  
'else': 0.004259134723156243,  
'incredibly': 0.0012889486662183366,  
'then': 0.02045505492042143,  
'suddenly': 0.0004203093476798924,  
'sad': 0.0015131136516476126,  
'mr': 0.00019614436225061645,  
'pink': 0.0007565568258238063,  
'genius': 0.0013169692893969962,  
'betcha': 2.8020623178659493e-05,  
'he': 0.0498767092580139,  
'explains': 0.0008966599417171038,  
'how': 0.024686169020399013,

'live': 0.007817753866845998,  
'strange': 0.0012609280430396772,  
'quasi-city': 2.8020623178659493e-05,  
'shadow': 0.0005604124635731899,  
'industry': 0.0012329074198610178,  
'curiously': 5.6041246357318986e-05,  
'full': 0.007313382649630128,  
'fun': 0.00829410446088321,  
'wasteland': 2.8020623178659493e-05,  
'other': 0.03104685048195472,  
'people': 0.018185384442950012,  
'dreams': 0.0005884330867518494,  
'photo': 0.004203093476798924,  
'video': 0.009078681909885675,  
'editing': 0.002325711723828738,  
'softwares': 2.8020623178659493e-05,  
'were': 0.022612642905178212,  
'basic': 0.002942165433759247,  
'old': 0.011908764850930285,  
'we': 0.02827280878726743,  
'upgrade': 0.0028861241874019276,  
'couple': 0.00498767092580139,  
'months': 0.0042871553463349026,  
'ago': 0.006556825823806321,  
'reading': 0.007817753866845998,  
'many': 0.02328513786146604,  
'reviews': 0.005015691548980049,  
'different': 0.010928043039677203,  
'bundles': 8.406186953597848e-05,  
'adobe': 0.004006949114548308,  
'photoshop': 0.003194351042367182,  
'elements': 0.0022416498542927594,  
'5.0': 0.00014010311589329746,  
'premiere': 0.0004763505940372114,  
'3.0': 8.406186953597848e-05,  
'bundle': 0.00036426810132257344,  
'played': 0.004091010984084286,  
'around': 0.009779197489352163,  
'bit': 0.00941492938802959,  
'awesome': 0.003698722259583053,  
'much': 0.028160726294552792,  
'pinnacle': 0.00025218560860793546,  
'studio': 0.0017372786370768885,  
'functions': 0.0011208249271463797,  
'intuitive': 0.0015411342748262722,  
'gone': 0.0022976911006500783,  
'figured': 0.0007285362026451469,  
'advanced': 0.0014570724052902937,  
'yet': 0.009190764402600313,  
'wonderful': 0.007901815736381977,  
'tool': 0.003110289172831204,  
'recently': 0.0030822685496525444,  
'scan': 0.0005604124635731899,  
'lot': 0.013702084734364493,  
'20+': 5.6041246357318986e-05,  
'year': 0.008938578793992378,  
'b': 0.001373010535754315,  
'w': 0.0011768661735036988,  
'photographs': 0.0007845774490024658,

'needed': 0.00439923783904954,  
'touching': 0.0009527011880744228,  
'up': 0.03684711947993723,  
'worked': 0.0037827841291190314,  
'wonderfully': 0.0008686393185384443,  
'could': 0.020623178659493386,  
'adjust': 0.0006164537099305089,  
'lighting': 0.0010367630576104013,  
'sharpening': 5.6041246357318986e-05,  
'fix': 0.0008966599417171038,  
'colors': 0.002045505492042143,  
'scratches': 0.00030822685496525445,  
'turned': 0.0018213405066128672,  
'amazing': 0.007005155794664874,  
'd': 0.006724949562878279,  
'never': 0.01597175521183591,  
'tell': 0.004511320331764179,  
'scanned': 0.0003922887245012329,  
'images': 0.004259134723156243,  
'such': 0.011320331764178435,  
'program': 0.013393857879399238,  
'things': 0.010647836807890608,  
'wanted': 0.0054079802734812825,  
'without': 0.013702084734364493,  
'having': 0.009835238735709483,  
'manual': 0.0038108047522976913,  
'surface': 0.0006164537099305089,  
'possible': 0.002325711723828738,  
'studying': 0.00036426810132257344,  
'detail': 0.0014290517821116341,  
'time': 0.038052006276619595,  
'need': 0.014682806545617575,  
'right': 0.010815960546962565,  
'now': 0.021491817978031832,  
'does': 0.0309908092355974,  
'there': 0.043235821564671595,  
'comment': 0.0008406186953597848,  
'why': 0.008966599417171038,  
'giving': 0.002101546738399462,  
'4': 0.00610849585294777,  
'stars': 0.00605245460659045,  
'5.': 0.0004483299708585519,  
'1.': 0.0011768661735036988,  
'minimum': 0.00028020623178659493,  
'requirements': 0.0003922887245012329,  
'512mb': 0.00011208249271463797,  
'ram': 0.0010928043039677203,  
'computer': 0.007565568258238063,  
'while': 0.017232683254875588,  
'painfully': 0.00019614436225061645,  
'slow': 0.003054247926473885,  
'feasible': 2.8020623178659493e-05,  
'ended': 0.0012889486662183366,  
'upgrading': 0.0005043712172158709,  
'1gb': 0.0005043712172158709,  
'been': 0.02874915938130464,  
'working': 0.0031663304191885226,  
'since': 0.01266532167675409,  
'make': 0.019026003138309797,

'sure': 0.007397444519166106,  
'software': 0.014206455951580363,  
'2.': 0.0013449899125756557,  
'red-eye': 0.00022416498542927594,  
'lack': 0.0016812373907195697,  
'thereof': 0.00014010311589329746,  
'auto': 0.001373010535754315,  
'finds': 0.0013449899125756557,  
'red': 0.002045505492042143,  
'eyes': 0.002549876709258014,  
'kinds': 0.0008125980721811253,  
'places': 0.001877381752970186,  
'where': 0.013477919748935216,  
'actually': 0.007677650750952701,  
'sometimes': 0.004175072853620265,  
'covers': 0.002045505492042143,  
'half': 0.0028020623178659494,  
'face': 0.0036707016364043938,  
'eye': 0.001709258013898229,  
'etc.': 0.004259134723156243,  
'option': 0.001877381752970186,  
'manually': 0.0004483299708585519,  
'maybe': 0.003194351042367182,  
'another': 0.013001569154898005,  
'trick': 0.0007005155794664874,  
'found': 0.01272136292311141,  
'usually': 0.0024658148397220356,  
'give': 0.009134723156242995,  
'shot': 0.003418516027796458,  
'take': 0.01325375476350594,  
'into': 0.023173055368751402,  
'disappointed': 0.002605917955615333,  
'fixing': 0.00019614436225061645,  
'something': 0.009302846895314952,  
'quite': 0.009078681909885675,  
'frequently': 0.0008966599417171038,  
'has': 0.056657700067249496,  
'amateur': 0.0006164537099305089,  
'o': 5.6041246357318986e-05,  
'recommendation': 0.0005323918403945303,  
'going': 0.009863259358888142,  
'definitely': 0.006584846446984981,  
'deserve': 0.0007565568258238063,  
'5': 0.008630351939027125,  
'points': 0.0017933198834342075,  
'james': 0.001933422999327505,  
'wrote': 0.0014570724052902937,  
'ben': 0.0003922887245012329,  
'weak': 0.0010367630576104013,  
'immature': 5.6041246357318986e-05,  
'stalker': 5.6041246357318986e-05,  
'lazy': 0.00025218560860793546,  
'bad': 0.007313382649630128,  
'taste': 0.0016812373907195697,  
'women': 0.003194351042367182,  
'21': 0.00028020623178659493,  
'unable': 0.0008125980721811253,  
'his': 0.05307106030038108,  
'parents': 0.0022136292311141,

'ask': 0.0015411342748262722,  
'him': 0.012076888590002242,  
'wear': 0.0008686393185384443,  
'scuba': 0.00016812373907195696,  
'suit': 0.0005604124635731899,  
'their': 0.03096278861241874,  
'friends': 0.003754763505940372,  
'wow': 0.0017372786370768885,  
'revolutionary': 0.00028020623178659493,  
'film-making...not': 2.8020623178659493e-05,  
'learning': 0.00495965030262273,  
'overcome': 0.00036426810132257344,  
'weakness': 0.00016812373907195696,  
'immaturity': 2.8020623178659493e-05,  
'defy': 5.6041246357318986e-05,  
'elaine': 0.00028020623178659493,  
'coming': 0.0022416498542927594,  
'age': 0.003222371665545842,  
'went': 0.004034969737726967,  
'over': 0.02163192109392513,  
'head': 0.003978928491369648,  
'acting': 0.002661959201972652,  
'dead': 0.002101546738399462,  
'subject': 0.002269670477471419,  
'matter': 0.003026227303295225,  
'shocker': 2.8020623178659493e-05,  
'came': 0.006977135171486214,  
'?': 0.020427034297242772,  
'grow': 0.0007285362026451469,  
'see': 0.018577673167451243,  
'value': 0.002830082941044609,  
'watch': 0.005464021519838601,  
'again': 0.01218897108271688,  
'few': 0.014178435328401703,  
'think': 0.01765299260255548,  
'classic': 0.00610849585294777,  
'used': 0.019110065007845775,  
'ef': 0.0005043712172158709,  
'75-300mm': 0.00011208249271463797,  
'4.5-5.6': 0.00014010311589329746,  
'iii': 0.00025218560860793546,  
'always': 0.012385115444967496,  
'noticed': 0.0014010311589329747,  
'limitations': 0.0005043712172158709,  
'shooting': 0.0026339385787939925,  
'softball': 5.6041246357318986e-05,  
'reviewed': 0.0005604124635731899,  
'lineup': 5.6041246357318986e-05,  
'website': 0.0019614436225061645,  
'100-400mm': 8.406186953597848e-05,  
'4.5-5.6l': 2.8020623178659493e-05,  
'add': 0.003418516027796458,  
'70-200mm': 0.00014010311589329746,  
'telephoto': 0.0006164537099305089,  
'collection': 0.00610849585294777,  
'serious': 0.002661959201972652,  
'meant': 0.0011488455503250391,  
'photographers': 0.0004763505940372114,  
'being': 0.013169692893969962,

'manufactured': 0.00016812373907195696,  
'materials': 0.0004483299708585519,  
'high': 0.007509527011880744,  
'level': 0.0032784129119031607,  
'dust': 0.0005884330867518494,  
'moisture': 0.0003922887245012329,  
'resistance': 0.0003362474781439139,  
'construction': 0.0008125980721811253,  
'sturdy': 0.0009246805648957633,  
'consists': 0.00025218560860793546,  
'23': 0.00022416498542927594,  
'18': 0.0004483299708585519,  
'groups': 0.0010087424344317419,  
'four': 0.004511320331764179,  
'ud-glass': 2.8020623178659493e-05,  
'largely': 0.0007285362026451469,  
'eliminate': 0.00030822685496525445,  
'secondary': 0.0003922887245012329,  
'spectrum': 0.0003922887245012329,  
'reducing': 0.0003362474781439139,  
'residual': 2.8020623178659493e-05,  
'chromatic': 2.8020623178659493e-05,  
'aberration': 2.8020623178659493e-05,  
'improved': 0.0016812373907195697,  
'8-blade': 2.8020623178659493e-05,  
'circular': 0.00025218560860793546,  
'aperture': 0.0005604124635731899,  
'offers': 0.0024097735933647163,  
'pleasing': 0.0003922887245012329,  
'out-of-focus': 5.6041246357318986e-05,  
'image': 0.006472763954270343,  
'weighs': 0.00019614436225061645,  
'3.24': 2.8020623178659493e-05,  
'lbs': 0.00025218560860793546,  
'1': 0.006584846446984981,  
'470': 2.8020623178659493e-05,  
'g': 0.0006444743331091683,  
'accepts': 0.00014010311589329746,  
'77mm': 2.8020623178659493e-05,  
'filters': 0.0009246805648957633,  
'minimal': 0.0004483299708585519,  
'focusing': 0.0004483299708585519,  
'distance': 0.0007845774490024658,  
'4.3': 2.8020623178659493e-05,  
'ft': 0.00022416498542927594,  
'1.3': 8.406186953597848e-05,  
'm.': 8.406186953597848e-05,  
'compatible': 0.0011768661735036988,  
'extenders': 2.8020623178659493e-05,  
'ii': 0.0016812373907195697,  
'2x': 0.00019614436225061645,  
'uv': 0.0004203093476798924,  
'haze': 0.00016812373907195696,  
'filter': 0.001597175521183591,  
'protect': 0.0010647836807890607,  
'front': 0.002381752970186057,  
'element': 0.0005884330867518494,  
'looks': 0.0035305985205110963,  
'helps': 0.00274602107150863,



'stay': 0.0024377942165433758,  
'cool': 0.0033344541582604795,  
'sun': 0.0014010311589329747,  
'some': 0.04004147052230442,  
'concerned': 0.0012048867966823582,  
'attracts': 8.406186953597848e-05,  
'attention': 0.0020735261152208025,  
'criminals': 0.00011208249271463797,  
'experience': 0.00498767092580139,  
'regardless': 0.0007285362026451469,  
'put': 0.00991930060524546,  
'addition': 0.002605917955615333,  
'most': 0.026227303295225286,  
'likely': 0.0012889486662183366,  
'tripod': 0.001989464245684824,  
'monopod': 0.0005323918403945303,  
'further': 0.0014010311589329747,  
'attract': 8.406186953597848e-05,  
'personally': 0.0015131136516476126,  
'doubt': 0.0015411342748262722,  
'thief': 0.00011208249271463797,  
'looking': 0.008966599417171038,  
'steal': 0.0005323918403945303,  
'taking': 0.005632145258910558,  
'evaluate': 5.6041246357318986e-05,  
'unload': 5.6041246357318986e-05,  
'thieves': 0.00016812373907195696,  
'insurance': 0.00014010311589329746,  
'dual': 0.0006724949562878278,  
'modes': 0.0005323918403945303,  
'mode': 0.0029982066801165657,  
'stationary': 0.00014010311589329746,  
'subjects': 0.0010647836807890607,  
'moving': 0.001933422999327505,  
'function': 0.0018213405066128672,  
';': 0.037267428827617126,  
'success': 0.0017372786370768885,  
'shake': 0.0009246805648957633,  
'stabilize': 5.6041246357318986e-05,  
'vehicle': 0.00014010311589329746,  
'extra': 0.0036426810132257344,  
'stops': 0.00036426810132257344,  
'mind': 0.00498767092580139,  
'attempts': 0.0008125980721811253,  
'movement': 0.0012889486662183366,  
'photographed': 0.00022416498542927594,  
'permits': 5.6041246357318986e-05,  
'focus': 0.0035586191436897557,  
'depending': 0.0005604124635731899,  
'switch': 0.0011488455503250391,  
'shift': 0.0006444743331091683,  
'1.4m': 2.8020623178659493e-05,  
'4.6ft': 2.8020623178659493e-05,  
'infinity': 0.00016812373907195696,  
'2.5m': 2.8020623178659493e-05,  
'8.2ft': 2.8020623178659493e-05,  
'included': 0.00386684599865501,  
'mount': 0.0005604124635731899,  
'storage': 0.0006164537099305089,

'mounting': 0.00028020623178659493,  
'balance': 0.0013449899125756557,  
'creates': 0.0007845774490024658,  
'unbalanced': 5.6041246357318986e-05,  
'may': 0.011824702981394306,  
'topple': 2.8020623178659493e-05,  
'zippered': 2.8020623178659493e-05,  
'lz1324': 2.8020623178659493e-05,  
'purchased': 0.0065288052006276616,  
'separately': 0.0003362474781439139,  
'et-86': 2.8020623178659493e-05,  
'reduce': 0.00030822685496525445,  
'flare': 0.0004203093476798924,  
'certain': 0.00165321676754091,  
'situations': 0.0013449899125756557,  
'often': 0.006024433983411791,  
'times': 0.010928043039677203,  
'recommends': 0.00028020623178659493,  
'purchasing': 0.0015411342748262722,  
'normally': 0.0007565568258238063,  
'almost': 0.008658372562205783,  
'fails': 0.00036426810132257344,  
'clarity': 0.0010087424344317419,  
'sharpness': 0.0008406186953597848,  
'refined': 0.00025218560860793546,  
'auto-focus': 0.00011208249271463797,  
'quick': 0.0036707016364043938,  
'portraits': 0.0004483299708585519,  
'landscape': 0.0006724949562878278,  
'sports': 0.0009527011880744228,  
'general': 0.0029141448105805875,  
'walk': 0.0014290517821116341,  
'2.8l': 8.406186953597848e-05,  
'light': 0.009106702533064335,  
'creative': 0.0013449899125756557,  
'versatility': 0.0003922887245012329,  
'comes': 0.0067809908092355976,  
'price': 0.012329074198610178,  
'worth': 0.010031383097960098,  
'incredible': 0.002157587984756781,  
'bulky': 0.0006164537099305089,  
'should': 0.013169692893969962,  
'expected': 0.002045505492042143,  
'types': 0.001373010535754315,  
'word': 0.004231114099977584,  
'user': 0.004315175969513562,  
'files': 0.0037827841291190314,  
'received': 0.0021856086079354406,  
'professional': 0.0044272584622282,  
'organizations': 0.00014010311589329746,  
'attorneys': 5.6041246357318986e-05,  
'microsoft': 0.003474557274153777,  
'works': 0.012525218560860793,  
'asset': 0.00016812373907195696,  
'business': 0.00386684599865501,  
'thank': 0.002381752970186057,  
'yo': 0.00030822685496525445,  
'invention': 8.406186953597848e-05,  
'warm': 0.0015691548980049316,

'smooth': 0.002101546738399462,  
'cream': 0.001709258013898229,  
'felt': 0.003026227303295225,  
'difference': 0.0024377942165433758,  
'cold': 0.0012609280430396772,  
'itself': 0.004091010984084286,  
'smoother': 0.0003362474781439139,  
'buttery': 5.6041246357318986e-05,  
'investment': 0.0010928043039677203,  
'space': 0.0024097735933647163,  
'language': 0.003054247926473885,  
'performed': 0.0008125980721811253,  
'designed': 0.0007845774490024658,  
'english': 0.0022416498542927594,  
'ok': 0.0017652992602555481,  
'razor': 0.0017933198834342075,  
'want': 0.01613987895090787,  
'photographing': 0.00014010311589329746,  
'wildlife': 0.0004203093476798924,  
'told': 0.002381752970186057,  
'fellow': 0.0007845774490024658,  
'film': 0.022864828513786146,  
'enthusiasts': 0.00016812373907195696,  
'bu\xlaue': 0.00022416498542927594,  
'later': 0.0045953822013001565,  
'films': 0.003978928491369648,  
'spanish': 0.002830082941044609,  
'master': 0.0013169692893969962,  
'earliest': 0.00014010311589329746,  
'films-': 5.6041246357318986e-05,  
'--his': 2.8020623178659493e-05,  
'famous': 0.0012889486662183366,  
'collaborations': 0.00014010311589329746,  
'salvador': 2.8020623178659493e-05,  
'dal\xla': 2.8020623178659493e-05,  
'instance-': 2.8020623178659493e-05,  
'--show': 2.8020623178659493e-05,  
'edgier': 2.8020623178659493e-05,  
'fascinating': 0.0014850930284689532,  
'whatever': 0.0017372786370768885,  
'saw': 0.00439923783904954,  
'1974': 0.00028020623178659493,  
'phantom': 0.00028020623178659493,  
'liberty': 0.00022416498542927594,  
'immediately': 0.0022416498542927594,  
'fell': 0.0009246805648957633,  
'who': 0.03799596503026227,  
'swear': 0.00016812373907195696,  
'popular': 0.0017652992602555481,  
'1972': 0.00011208249271463797,  
'oscar-winner': 2.8020623178659493e-05,  
'discreet': 0.00014010311589329746,  
'charm': 0.0003362474781439139,  
'bourgeoisie': 2.8020623178659493e-05,  
'somehow': 0.0009807218112530822,  
'entertaining': 0.0017933198834342075,  
'plot': 0.00330643353508182,  
'speak': 0.0015131136516476126,  
'basically': 0.0016812373907195697,

```

'surrealist': 2.8020623178659493e-05,
'sketches': 0.00022416498542927594,
'playing': 0.0045673615781214975,
'ideas': 0.001877381752970186,
'monks': 0.00011208249271463797,
'pray': 0.00019614436225061645,
'woman': 0.0029701860569379063,
'sick': 0.0008406186953597848,
'father': 0.0029141448105805875,
'poker': 5.6041246357318986e-05,
'smoke': 0.00019614436225061645,
'group': 0.003138309796009863,
'sit': 0.0015131136516476126,
'dinner': 0.0005323918403945303,
'table': 0.0008125980721811253,
'toilets': 0.00011208249271463797,
'go': 0.0154954046177987,
'bathroom': 0.0006164537099305089,
'eat': 0.0016812373907195697,
'desperately': 0.00019614436225061645,
'missing': 0.0014850930284689532,
'daughter-': 5.6041246357318986e-05,
'---even': 2.8020623178659493e-05,
'class': 0.0022416498542927594,
'call': 0.0022136292311141,
'name': 0.0032503922887245013,
'universe': 0.0006444743331091683,
'concocts': 2.8020623178659493e-05,
'goes': 0.004147052230441605,
'thing': 0.010115444967496078,
'seeming': 0.00014010311589329746,
'irrational': 8.406186953597848e-05,
'surrealistic': 5.6041246357318986e-05,
'broad': 0.00030822685496525445,
'theme': 0.0015411342748262722,
'support': 0.0029701860569379063,
'free-form': 5.6041246357318986e-05,
'structure': 0.0010647836807890607,
'comic': 0.0012329074198610178,
'vision': 0.0007285362026451469,
'freedom': 0.0007005155794664874,
'run': 0.0045673615781214975,
'amuck': 5.6041246357318986e-05,
'idea': 0.0029141448105805875,
'appealing': 0.0004763505940372114,
'everyone...but': 2.8020623178659493e-05,
'suggesting': 5.6041246357318986e-05,
'limits': 0.00036426810132257344,
'opening': 0.0017933198834342075,
'scene': 0.0034465366509751176,
'toledo': 2.8020623178659493e-05,
'spain': 0.00025218560860793546,
'1808': 2.8020623178659493e-05,
'napoleon': 5.6041246357318986e-05,
...}}

```

## Classifying new documents

```

In [ ]: # Calculates the score of a document given a label and a trained naive bayes
def score_doc_label(document, label, word_count, word_probs) -> float:
    # Adding smoothing -> 1/(total words for label)
    word_scores = []
    total_words = word_count[label]
    for word in document:
        if word not in word_probs[label].keys():
            word_scores.append(1/total_words)
        else:
            word_scores.append(word_probs[label][word])
    # Previous score
    #doc_score = functools.reduce(lambda x, y: x*y, word_scores)
    likelihoods = functools.reduce(lambda x, y: x*y, word_scores)
    prior = total_words/sum(word_count.values())
    doc_score = likelihoods*prior

    return doc_score

word_count = count_total_words(count_words(train_docs, train_labels))
res = score_doc_label(['a', 'top-quality', 'performance'], 'neg', word_count)

# Classifies the document given a trained naive bayes model
# Returns a tuple containing the classification and the score
def classify_nb(document, word_count, word_probs) -> tuple:
    result_label = ''
    # Set to smallest possible value, such that log probs works
    maximum = float('-inf')
    for label in word_probs.keys():
        score = score_doc_label(document, label, word_count, word_probs)
        if score >= maximum:
            maximum = score
            result_label = label
    return (result_label, maximum)

(result, score) = classify_nb(['a', 'top-quality', 'performance'], word_count)
result

```

Out [ ]: 'pos'

## Evaluating the classifier

```

In [ ]: # Classifies a set of documents (validation data) from a trained naive bayes
def classify_documents(docs, word_count, word_probs) -> list:
    results = []
    for doc in docs:
        # Appends the label of the classification
        results.append(classify_nb(doc, word_count, word_probs)[0])
    return results

# Computes the accuracy of a given classification result
def accuracy(true_labels, predicted_labels) -> float:
    correct = 0
    total = len(true_labels)
    for i in range(total):
        if true_labels[i] == predicted_labels[i]:
            correct += 1
    return correct / total

# Can be used to calculate true positives, true negatives,
# false positives and false negatives
def calc_metric(true_labels, predicted_labels, label, bin_op1, bin_op2) -> int:
    res = 0
    for i in range(len(true_labels)):
        comparison1 = bin_op1(true_labels[i], label)
        comparison2 = bin_op2(predicted_labels[i], label)
        if comparison1 and comparison2:
            res += 1
    return res

# Calculates the f1 score given true outcomes, predicted outcomes and true label
def calc_f1_score(t_labels, p_labels, t_label) -> float:
    true = lambda x, y: x==y
    false = lambda x, y: x!=y
    tp = calc_metric(t_labels, p_labels, t_label, true, true)
    fn = calc_metric(t_labels, p_labels, t_label, true, false)
    fp = calc_metric(t_labels, p_labels, t_label, false, true)

    # To avoid division by zero, in this case just return 0
    if ((fp+tp) == 0 or tp+fn == 0):
        return 0
    precision = tp / (fp+tp)
    recall = tp / (tp+fn)
    score = 2 * ((precision*recall) / (precision+recall))
    return score

predictions = classify_documents(val_docs, word_count, train)
print('Accuracy:', accuracy(val_labels, predictions))
print('Computed f1_score:', calc_f1_score(val_labels, predictions, 'pos'))

```

```

Accuracy: 0.7058329836340747
Computed f1_score: 0.6880284824210057
Scikit-learn f1_score: 0.6880284824210057

```

## Discussion about the differences between accuracy score and f1 score

*Accuracy* is easy to interpret, the statistic is obvious in what it stands for; x% accuracy means that x% was correctly classified. But it lacks in taking the distribution into consideration.

This is where the *f1-score* compliments the accuracy score. The f1-score will provide a better assesment of inbalanced data, it also considers the impact of having a false negative, which in many cases is highly relevant, for instance in medical measurments. The f1-score could be viewed as more strict and a more broad measurment, however, it is harder to interpret than the accuracy.

## Error analysis

```
In [ ]: # Finds the miss-classified documents and stores relevant information about
def find_miss_classifications(true_labels, val_docs, nb) -> list:
    result = []
    word_count = count_total_words(count_words(val_docs, true_labels))
    for i in range(len(true_labels)):
        (label, score) = classify_nb(val_docs[i], word_count, nb)
        if true_labels[i] != label:
            result.append({
                'Expected Label': true_labels[i],
                'Predicted label': label,
                'Score': np.log(score),
                'Document': val_docs[i]
            })

    return result

# Finds the worst classifications, the ones who are wrong and have a relative
# given the amount to output and the classifications
def find_worst_predictions(amount, predictions) -> list:
    worst_predictions = []
    for prediction in predictions:
        if len(worst_predictions) >= amount:
            # Sort the list of dictionaries by score, lowest last (since it
            worst_predictions.sort(key=lambda x: x['Score'], reverse=True)
            if prediction['Score'] > worst_predictions[-1]['Score']:
                # Remove the best and replace it with a worse one
                worst_predictions.pop()
                worst_predictions.append(prediction)
        else:
            worst_predictions.append(prediction)
    return worst_predictions

miss_classifications = find_miss_classifications(val_labels, val_docs, train
worst_predictions = find_worst_predictions(5, miss_classifications)
for i in range(len(worst_predictions)):
    print(f'{i + 1}. \n{worst_predictions[i]} \n')
```

```

1.
{'Expected Label': 'pos', 'Predicted label': 'neg', 'Score': -9.33520935402
2052, 'Document': ['goo']}

2.
{'Expected Label': 'neg', 'Predicted label': 'pos', 'Score': -20.0542150500
28432, 'Document': ['it', "'s", 'not', 'great', 'music', 'to', 'dance', 'to
']}

3.
{'Expected Label': 'pos', 'Predicted label': 'neg', 'Score': -45.0963240089
38005, 'Document': ['this', 'was', 'not', 'a', 'waste', 'of', 'brad', "'s",
'time', 'or', 'mine', '.', 'good', 'job', '!']}

4.
{'Expected Label': 'neg', 'Predicted label': 'pos', 'Score': -47.0609362304
16246, 'Document': ['this', 'lifts', 'your', 'heart', 'to', 'god', 'as', 'y
ou', 'listen', '.', 'beautifully', 'done']}

5.
{'Expected Label': 'pos', 'Predicted label': 'neg', 'Score': -48.9116508612
9675, 'Document': ['you', 'will', 'never', 'look', 'at', 'these', 'birds',
'the', 'same', 'again', '-', 'fansinating']}

```

```

/var/folders/st/ztz092hj7sdcnz4hsgfxjgqh0000gn/T/ipykernel_71899/427402592
3.py:11: RuntimeWarning: divide by zero encountered in log
  'Score': np.log(score),

```

### Notes on the error analysis

From the functions above and its output, we can identify the five worst classifications, i.e. the miss-classifications with highest score, from top to bottom. The scores are displayed on a logarithmic scale.

The worst classification is in some sense not relevant, since it shouldn't be viewed as a meaningful review, given its only one word and not a real word at that. The relative high score is due to the few words. The other ones show that the model isn't ideal, however, one could argue that they're not obvious if you view them as separate words. This points at the flaw of the Naive Bayes model; no context is taken into consideration, every word is viewed as independent. This leads to some seemingly very obvious classifications of the reviews resulting in miss-classifications. I think that the fifth review above is a good example of a review that's pretty obviously positive, but no words in isolation indicates that this is more of a positive review than a negative one, thus the classification could either be negative or positive, in our case negative. If some type of context was considered, in a more involved model, the fifth review would probably not be miss-classified.

### Cross-validation

#### 10-fold cross validation



```
In [ ]: '''
TODO for N iterations, in our case 10, split the in two distinct chunks 10 t
check the score for each result and combine them.
It's only going to validate a chunk at a time, so when the iteration is done
combine all the results and validate it for all the true labels, not only th
'''

def n_fold_cross_validation(n: int, docs: list, labels: list):
    # To store all the predictions
    results = []

    for fold_nbr in range(n):
        split_point_1 = int(float(fold_nbr)/n*len(docs))
        split_point_2 = int(float(fold_nbr+1)/n*len(docs))

        train_docs_fold = docs[:split_point_1] + docs[split_point_2:]
        train_labels_fold = labels[:split_point_2] + labels[split_point_2:]
        val_docs_fold = docs[split_point_1:split_point_2]

        # Train the classifier
        word_probs = train_nb(train_docs_fold, train_labels_fold)
        word_count = count_total_words(count_words(train_docs_fold, train_la
        # Apply the classifier to val_docs_fold
        predictions = classify_documents(val_docs_fold, word_count, word_pro
        results.append(predictions)

    # Flatten the array
    results = sum(results, [])

    return results

# 10 fold cross validation
predictions = n_fold_cross_validation(10, all_docs, all_labels)
print('Metrics for 10 fold cross validation')
print('Accuracy:', accuracy(all_labels, predictions))
print('Computed f1_score:', calc_f1_score(all_labels, predictions, 'pos'))
print('Scikit-learn f1_score:', f1_score(all_labels, predictions, pos_label=

Metrics for 10 fold cross validation
Accuracy: 0.6432768171898606
Computed f1_score: 0.6791484221651819
Scikit-learn f1_score: 0.6791484221651819
```

### Leave-one-out cross validation

```
In [ ]: # Leave-one-out cross validation
# This one takes approximately 52 minutes to compute on my hardware
predictions = n_fold_cross_validation(len(all_docs), all_docs, all_labels)
print('Metrics for Leave-one-out cross validation')
print('Accuracy:', accuracy(all_labels, predictions))
print('Computed f1_score:', calc_f1_score(all_labels, predictions, 'pos'))
print('Scikit-learn f1_score:', f1_score(all_labels, predictions, pos_label=
```

```
11914
11914
Metrics for Leave-one-out cross validation
Accuracy: 0.6569582004364614
Computed f1_score: 0.651547446500128
Scikit-learn f1_score: 0.651547446500128
```

### **Difference between 10-fold cross validation and Leave-one-out cross validation**

The difference between the two validation methods lies mainly in how big the data set is. In a world with endlessly fast computational methods Leave-one-out cross validation would almost always be preferred, since it trains the data on every single data point, but more importantly tests each data point one at a time. However, for very large data sets, using Leave-one-out cross validation isn't realistic and perhaps not even necessary given the time it takes and the accuracy one can achieve using only k-fold cross validation. In our case, 10-fold cross validation is a sufficient validation metric, since it gives a very close score to Leave-one-out cross validation, in a fraction of the time. In the model for the Iris data set however, Leave-one-out cross validation could be preferred, since it wouldn't take much time to run it, due to the relatively small data set (150 data points) and the fact that it validates each and every data point.

### **Naive Bayes for numerical data**

```

In [ ]: flower_data = pd.read_csv("iris.csv")
flower_X = flower_data.iloc[:, :-1].to_numpy() #flower_data.drop('species', a
flower_Y = flower_data['species'].to_numpy()

'''
Then when you predict you assume that they follow a gaussian distribution, w
Features: sepal length (x1), sepal_width (x2), petal_length (x3), petal_wid
Likelihood:  $P(X = [x_1, x_2, x_3, x_4] \mid Y = y)$  (uses PDF)
Prior:  $P(Y = y)$ 
'''
def train_gnb(X, labels) -> dict:
    label_freqs = Counter(labels)
    label_values = {}
    vector_len = len(X[0])
    i = 0

    # Compute statistic for each X for each label
    for label in label_freqs.keys():
        frequency = label_freqs[label]
        # Split X into the chunk of the specific label
        label_X = X[i:i+frequency]
        label_values.update({label: {}})
        # Compute statistic for each column for each label
        for j in range(vector_len):
            x_mean = np.mean(label_X[j])
            x_std = np.std(label_X[j])
            label_values[label].update({f'x{j}': {
                'mean': x_mean,
                'std': x_std
            }})
        i += frequency

    return label_values

def score_label(values, label, frequencies, model):
    likelihoods = []
    for i in range(len(values)):
        mean = model[label][f'x{i}']['mean']
        std = model[label][f'x{i}']['std']
        #This could be incorrect
        likelihood = (1 / (np.sqrt(2 * np.pi) * std)) * np.exp(-((values[i]
        likelihoods.append(likelihood)

    prior = frequencies[label] / sum(frequencies.values())
    score = np.log(functools.reduce(lambda x, y: x*y, likelihoods) * prior)

    return score

def classify_gnb(x, frequencies, model):
    result_label = ''
    maximum = float('-inf')
    for label in model.keys():
        score = score_label(x, label, frequencies, model)
        if score >= maximum:
            maximum = score
            result_label = label
    return (result_label, score)

```

```
# Classifies iris data (validation data) from a trained Gaussian naive bayes
def classify_all(X, frequencies, model) -> list:
    results = []
    for x in X:
        # Appends the label of the classification
        results.append(classify_gnb(x, frequencies, model)[0])
    return results
```

```
r = train_gnb(flower_X, flower_Y)
print(r)
classify_gnb([5.1,3.5,1.4,0.2], Counter(flower_Y), r)
```

```
{'setosa': {'x0': {'mean': 2.55, 'std': 1.8874586088176872}, 'x1': {'mean': 2.375, 'std': 1.7640507362318127}, 'x2': {'mean': 2.35, 'std': 1.7298843892006195}, 'x3': {'mean': 2.3499999999999996, 'std': 1.6560495161679194}}, 'versicolor': {'x0': {'mean': 4.074999999999999, 'std': 2.0535031044534606}, 'x1': {'mean': 3.9000000000000004, 'std': 1.7930421077041108}, 'x2': {'mean': 4.1, 'std': 2.0149441679609885}, 'x3': {'mean': 3.2750000000000004, 'std': 1.6068213964221414}}, 'virginica': {'x0': {'mean': 4.525, 'std': 1.6528384676065595}, 'x1': {'mean': 3.875, 'std': 1.6192204914711275}, 'x2': {'mean': 4.525, 'std': 2.044963324854507}, 'x3': {'mean': 4.1499999999999995, 'std': 1.8580904176062045}}}
```

Out[ ]: ('setosa', -10.608272345678214)

```

In [ ]: # Validating the gaussian nb model
def n_fold_cross_validation_nb(N: int, data):
    # To store all the predictions
    results = []
    #print(random.shuffle(data.to_numpy()))

    docs = list(data.iloc[:, :-1].to_numpy())
    labels = list(data['species'].to_numpy())

    #print(pd.DataFrame(random.shuffle(data.to_numpy())))

    for fold_nbr in range(N):
        split_point_1 = int(float(fold_nbr)/N*len(docs))
        split_point_2 = int(float(fold_nbr+1)/N*len(docs))

        train_docs_fold = docs[:split_point_1] + docs[split_point_2:]
        train_labels_fold = labels[:split_point_1] + labels[split_point_2:]
        val_docs_fold = docs[split_point_1:split_point_2]

        # Train the classifier
        model = train_gnb(train_docs_fold, train_labels_fold)
        frequencies = Counter(labels)
        # Apply the classifier to val_docs_fold
        predictions = classify_all(val_docs_fold, frequencies, model)
        results.append(predictions)

    # Flatten the array
    results = sum(results, [])

    return results

#flower_data = flower_data.sample(frac=1)

predictions = n_fold_cross_validation_nb(10, flower_data)
print('Metrics for 10 fold cross validation')
print('Accuracy:', accuracy(flower_Y, predictions))
print('F1_score for setosa:', calc_f1_score(flower_Y, predictions, 'setosa'))
print('F1_score for versicolor:', calc_f1_score(flower_Y, predictions, 'versicolor'))
print('F1_score for virginica:', calc_f1_score(flower_Y, predictions, 'virginica'))

```

Metrics for 10 fold cross validation

Accuracy: 0.6

F1\_score for setosa: 0.9523809523809523

F1\_score for versicolor: 0.5384615384615384

F1\_score for virginica: 0.15384615384615383

**Discussion about the results** As shown by the f1-scores the model performs very well on 'setosa', worse on 'versicolor' and very bad on 'virginica'. This is probably due to some error leading the model to start with great predictions, but increasingly bad when the category jumps from the first to the second, and the second to the third. I suspect that it has something to do with the data being ordered, and the model being biased to the order in which it classifies, but I tried to randomise the order and it didn't produce a better result, so I suspect that I've missed something in my implementation of the model. Unfortunately, due to the time, I wasn't able to find this error. That being said, the model works well initially, but has a drastic decrease in performance when trying to classify the other two species