

Author Classification

import dependences

```
In [2]: import os
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
stopwords = stopwords.words('english')
```

read data

```
In [3]: texts = pd.read_csv('federalist.csv')
texts
```

```
Out[3]:
```

	author	text
0	HAMILTON	FEDERALIST. No. 1 General Introduction For the...
1	JAY	FEDERALIST No. 2 Concerning Dangers from Forei...
2	JAY	FEDERALIST No. 3 The Same Subject Continued (C...
3	JAY	FEDERALIST No. 4 The Same Subject Continued (C...
4	JAY	FEDERALIST No. 5 The Same Subject Continued (C...
...
78	HAMILTON	FEDERALIST No. 79 The Judiciary Continued From...
79	HAMILTON	FEDERALIST No. 80 The Powers of the Judiciary ...
80	HAMILTON	FEDERALIST. No. 81 The Judiciary Continued, an...
81	HAMILTON	FEDERALIST No. 82 The Judiciary Continued From...
82	HAMILTON	FEDERALIST No. 83 The Judiciary Continued in R...

83 rows × 2 columns

```
In [4]: # convert author colum to label column
authors = set(texts['author'])
authors = {author : idx for author, idx in zip(authors, range(len(authors)))}
texts['author'] = texts['author'].map(lambda x: authors[x])
texts.head()
```

Out[4]:

	author	text
0	3	FEDERALIST. No. 1 General Introduction For the...
1	4	FEDERALIST No. 2 Concerning Dangers from Forei...
2	4	FEDERALIST No. 3 The Same Subject Continued (C...
3	4	FEDERALIST No. 4 The Same Subject Continued (C...
4	4	FEDERALIST No. 5 The Same Subject Continued (C...

In [5]: `# count documents by author`
`texts['author'].value_counts()`

Out[5]:

3	49
0	15
1	11
4	5
2	3

Name: author, dtype: int64

Split data into training and validation sets

In [6]: `train_texts, val_texts, train_labels, val_labels = train_test_split(texts['te`

In [7]: `print("Traing text shape", train_texts.shape)`
`print("Validataion text shape", val_texts.shape)`

Traing text shape (66,)
 Validataion text shape (17,)

In [9]: `print("Traing label shape", train_labels.shape)`
`print("Validataion label shape", val_labels.shape)`

Traing label shape (66,)
 Validataion label shape (17,)

Text Processing

In [168... `# initializer tf-idf vectorizer`
`from sklearn.feature_extraction.text import TfidfVectorizer`
`vectorizer = TfidfVectorizer(stop_words = stopwords)`
`vectorizer.fit(train_texts)`

Out[168... `TfidfVectorizer(stop_words=['i', 'me', 'my', 'myself', 'we', 'our', 'ours',`
`'ourselves', 'you', "you're", "you've", "you'll",`
`"you'd", 'your', 'yours', 'yourself', 'yourselves'`
`,`
`'he', 'him', 'his', 'himself', 'she', "she's",`
`'her', 'hers', 'herself', 'it', "it's", 'its',`
`'itself', ...])`

```
In [169... train_X = vectorizer.transform(train_texts)
            val_X = vectorizer.transform(val_texts)
```

Bernoulli Naïve Bayes

```
In [56]: from sklearn.naive_bayes import BernoulliNB
         nv = BernoulliNB()
```

```
In [57]: # train Naïve Bayes
         nv.fit(train_X, train_labels)
```

```
Out[57]: BernoulliNB()
```

```
In [58]: # evaluate Naïve Bayes

         # with training set
         print("Training accuracy", nv.score(train_X, train_labels))

         # with validation set
         print("Training accuracy", nv.score(val_X, val_labels))
```

```
Training accuracy 0.7727272727272727
Training accuracy 0.5882352941176471
```

```
In [60]: len(vectorizer.vocabulary_)
```

```
Out[60]: 7876
```

Reduce max-features to 1000

```
In [299... vectorizer = TfidfVectorizer(stop_words = stopwords,
                                max_features = 1000,
                                ngram_range = (1,2))
         vectorizer.fit(train_texts)
```

```
Out[299... TfidfVectorizer(max_features=1000,
                        stop_words=['i', 'me', 'my', 'myself', 'we', 'our', 'ours',
                                   'ourselves', 'you', "you're", "you've", "you'll",
                                   "you'd", 'your', 'yours', 'yourself', 'yourselves',
                                   'he', 'him', 'his', 'himself', 'she', "she's",
                                   'her', 'hers', 'herself', 'it', "it's", 'its',
                                   'itself', ...])
```

```
In [300... train_X = vectorizer.transform(train_texts)
            val_X = vectorizer.transform(val_texts)
```

```
In [301... vectorizer.vocabulary_
```

```
Out[301... {'federalist': 364,
```

'executive': 331,
'department': 231,
'new': 599,
'york': 998,
'packet': 641,
'tuesday': 935,
'1788': 0,
'people': 659,
'state': 864,
'constitution': 187,
'proposed': 725,
'government': 409,
'claims': 132,
'next': 600,
'attention': 79,
'hardly': 426,
'part': 644,
'system': 890,
'could': 202,
'attended': 78,
'greater': 415,
'difficulty': 244,
'perhaps': 661,
'less': 520,
'judgment': 484,
'seem': 821,
'taken': 892,
'upon': 947,
'opposition': 632,
'intended': 473,
'president': 697,
'united': 943,
'states': 866,
'merely': 565,
'full': 397,
'establish': 305,
'resources': 789,
'even': 311,
'authorities': 80,
'magistrate': 540,
'instances': 469,
'governor': 411,
'superior': 881,
'dignity': 245,
'king': 495,
'great': 414,
'britain': 97,
'shown': 839,
'us': 949,
'foreign': 380,
'scarcely': 814,
'future': 401,
'attempts': 76,
'might': 568,
'rather': 751,
'said': 808,
'object': 607,

'render': 774,
'necessary': 591,
'take': 891,
'view': 962,
'real': 754,
'nature': 588,
'form': 381,
'order': 634,
'well': 977,
'true': 932,
'appearance': 53,
'execution': 330,
'task': 894,
'man': 544,
'would': 994,
'find': 370,
'either': 279,
'public': 738,
'opinion': 629,
'relation': 768,
'subject': 871,
'far': 359,
'exceed': 324,
'usual': 951,
'though': 911,
'party': 650,
'disposition': 252,
'candid': 104,
'must': 581,
'force': 378,
'sentiments': 828,
'favor': 361,
'construction': 190,
'conduct': 163,
'political': 675,
'adversaries': 24,
'give': 405,
'place': 667,
'impossible': 450,
'character': 120,
'still': 867,
'employed': 285,
'success': 877,
'one': 625,
'instance': 468,
'general': 402,
'spirit': 859,
'proceeded': 714,
'power': 684,
'expressly': 346,
'individual': 460,
'mean': 555,
'vacancies': 956,
'senate': 824,
'experiment': 342,
'whatever': 978,
'may': 554,

'merit': 566,
'share': 836,
'observations': 613,
'equally': 300,
'let': 521,
'evidence': 316,
'fact': 353,
'able': 1,
'justify': 490,
'truth': 934,
'rules': 804,
'fair': 357,
'second': 817,
'clause': 135,
'article': 70,
'advice': 25,
'consent': 174,
'ambassadors': 40,
'ministers': 573,
'consuls': 191,
'judges': 483,
'supreme': 889,
'court': 209,
'officers': 621,
'whose': 984,
'appointments': 62,
'otherwise': 639,
'provided': 733,
'shall': 835,
'established': 306,
'law': 507,
'immediately': 443,
'another': 48,
'words': 991,
'happen': 423,
'end': 288,
'last': 504,
'provision': 735,
'connection': 173,
'obvious': 617,
'meaning': 556,
'terms': 902,
'first': 371,
'two': 938,
'clear': 136,
'mode': 575,
'appointing': 60,
'course': 208,
'cannot': 105,
'extend': 347,
'senators': 825,
'require': 782,
'establishment': 308,
'position': 679,
'understood': 940,
'following': 377,
'reasons': 757,

'nothing': 602,
'purpose': 741,
'establishing': 307,
'method': 567,
'appointment': 61,
'cases': 113,
'ordinary': 635,
'confined': 169,
'therefore': 905,
'exercised': 333,
'improper': 452,
'body': 93,
'continually': 194,
'service': 833,
'without': 990,
'evidently': 318,
'make': 542,
'temporary': 898,
'considered': 182,
'preceding': 689,
'seen': 823,
'description': 237,
'members': 561,
'time': 918,
'within': 989,
'operate': 626,
'duration': 269,
'sense': 826,
'naturally': 587,
'legislatures': 519,
'permanent': 663,
'national': 584,
'extended': 348,
'office': 620,
'legislature': 518,
'representation': 776,
'happened': 424,
'instead': 470,
'making': 543,
'circumstances': 126,
'authorized': 82,
'governed': 408,
'situation': 845,
'alone': 33,
'examination': 320,
'founded': 390,
'respect': 790,
'concurrent': 161,
'agency': 28,
'third': 909,
'doubt': 265,
'pretext': 700,
'former': 384,
'composed': 155,
'chosen': 124,
'six': 846,
'years': 996,

'latter': 506,
'express': 345,
'given': 406,
'supposition': 888,
'strong': 869,
'light': 527,
'proof': 720,
'arts': 72,
'prevent': 702,
'impartial': 444,
'submitted': 873,
'consideration': 180,
'case': 112,
'little': 533,
'papers': 643,
'decision': 219,
'whether': 980,
'language': 501,
'furnish': 400,
'much': 580,
'attempt': 75,
'citizens': 129,
'america': 42,
'publius': 739,
'see': 820,
'total': 921,
'number': 604,
'house': 434,
'representatives': 778,
'friday': 395,
'consist': 183,
'forms': 386,
'point': 672,
'branch': 95,
'federal': 363,
'indeed': 456,
'whole': 982,
'seems': 822,
'rendered': 775,
'weight': 976,
'argument': 64,
'small': 848,
'interests': 476,
'possess': 680,
'proper': 721,
'knowledge': 497,
'local': 534,
'numerous': 606,
'constituents': 185,
'class': 133,
'least': 511,
'mass': 548,
'likely': 529,
'many': 547,
'increase': 454,
'remarked': 772,
'precise': 690,

'representative': 777,
'policy': 674,
'several': 834,
'legislative': 517,
'directly': 248,
'consider': 178,
'bear': 84,
'difference': 240,
'consists': 184,
'twenty': 937,
'massachusetts': 549,
'three': 914,
'four': 391,
'hundred': 437,
'considerable': 179,
'among': 44,
'nearly': 589,
'equal': 298,
'population': 677,
'pennsylvania': 658,
'mentioned': 563,
'carolina': 110,
'five': 373,
'island': 479,
'proportion': 724,
'every': 315,
'thousand': 913,
'according': 4,
'carried': 111,
'remark': 771,
'made': 538,
'ought': 640,
'regulated': 764,
'standard': 862,
'amount': 45,
'hence': 429,
'hand': 421,
'applied': 57,
'assembly': 74,
'found': 388,
'principles': 707,
'men': 562,
'properly': 722,
'degree': 227,
'follow': 376,
'better': 89,
'reasoning': 756,
'certain': 117,
'secure': 818,
'free': 392,
'discussion': 250,
'guard': 419,
'easy': 273,
'combination': 142,
'purposes': 742,
'kept': 493,
'avoid': 83,

'passion': 653,
'never': 598,
'reason': 755,
'citizen': 128,
'also': 35,
'elections': 282,
'limited': 531,
'powers': 686,
'congress': 170,
'control': 198,
'frequent': 393,
'safely': 806,
'need': 594,
'possessed': 681,
'legislation': 516,
'bodies': 92,
'mind': 571,
'objections': 609,
'stated': 865,
'inhabitants': 466,
'period': 662,
'continue': 195,
'limitation': 530,
'thought': 912,
'raise': 748,
'doubted': 266,
'already': 34,
'presume': 698,
'put': 744,
'arising': 66,
'granted': 413,
'objection': 608,
'show': 838,
'manner': 546,
'contrary': 197,
'admit': 20,
'question': 746,
'decided': 218,
'regulation': 766,
'dangerous': 215,
'liberty': 524,
'negative': 595,
'answer': 49,
'regard': 760,
'present': 694,
'genius': 404,
'year': 995,
'choice': 123,
'disposed': 251,
'scheme': 815,
'tyranny': 939,
'feel': 366,
'motives': 579,
'means': 557,
'fail': 356,
'liberties': 523,
'common': 148,

'short': 837,
'capable': 106,
'large': 502,
'trust': 933,
'committed': 147,
'change': 119,
'country': 207,
'produce': 716,
'requires': 784,
'declare': 221,
'pretensions': 699,
'probable': 711,
'hands': 422,
'danger': 214,
'proceed': 713,
'easily': 272,
'rulers': 803,
'enable': 286,
'independent': 458,
'nation': 583,
'revolution': 798,
'appointed': 59,
'pleasure': 671,
'generally': 403,
'continued': 196,
'articles': 71,
'longer': 536,
'term': 901,
'always': 39,
'sole': 851,
'affairs': 26,
'nations': 585,
'war': 970,
'ever': 314,
'supposed': 887,
'use': 950,
'yet': 997,
'experience': 341,
'councils': 204,
'particular': 646,
'branches': 96,
'previous': 703,
'different': 241,
'private': 708,
'american': 43,
'sources': 855,
'sometimes': 852,
'virtue': 966,
'standing': 863,
'republican': 780,
'society': 850,
'placed': 668,
'civil': 131,
'offices': 622,
'increased': 455,
'election': 281,
'existing': 337,

'become': 85,
'suppose': 886,
'sufficient': 878,
'rule': 802,
'events': 313,
'calculated': 101,
'substitute': 876,
'jealousy': 480,
'friends': 396,
'cause': 114,
'mankind': 545,
'human': 436,
'portion': 678,
'confidence': 168,
'existence': 336,
'drawn': 267,
'inference': 462,
'characters': 121,
'plan': 670,
'convention': 200,
'serve': 832,
'thing': 906,
'authority': 81,
'exceptions': 326,
'vested': 961,
'single': 844,
'however': 435,
'comparison': 152,
'elected': 280,
'often': 623,
'think': 908,
'hereditary': 430,
'monarch': 577,
'requisite': 785,
'influence': 464,
'like': 528,
'throughout': 915,
'chief': 122,
'union': 942,
'liable': 522,
'tried': 930,
'conviction': 201,
'punishment': 740,
'person': 664,
'constitutional': 188,
'important': 447,
'circumstance': 125,
'personal': 665,
'responsibility': 793,
'stand': 861,
'ground': 418,
'passed': 652,
'absolute': 2,
'acts': 9,
'wholly': 983,
'majority': 541,
'necessity': 593,

'qualified': 745,
'british': 98,
'sovereign': 856,
'council': 203,
'original': 637,
'army': 69,
'militia': 570,
'called': 103,
'actual': 10,
'except': 325,
'measures': 559,
'judge': 482,
'expedient': 339,
'occasions': 619,
'care': 109,
'laws': 508,
'executed': 329,
'material': 550,
'points': 673,
'command': 144,
'times': 919,
'entire': 294,
'inferior': 463,
'substance': 875,
'direction': 247,
'military': 569,
'forces': 379,
'confederacy': 165,
'regulating': 765,
'fleets': 375,
'armies': 67,
'constitutions': 189,
'larger': 503,
'respective': 791,
'consequences': 176,
'kind': 494,
'head': 428,
'effects': 275,
'prospect': 728,
'mere': 564,
'final': 368,
'appeal': 50,
'arms': 68,
'probability': 710,
'afford': 27,
'measure': 558,
'matter': 551,
'enemies': 289,
'aid': 30,
'similar': 841,
'treaties': 926,
'peace': 656,
'commerce': 145,
'conclusive': 159,
'believe': 86,
'doctrine': 262,
'occasion': 618,

'kingdom': 496,
'complete': 154,
'legal': 515,
'sanction': 810,
'treaty': 927,
'co': 138,
'operation': 627,
'efficacy': 277,
'revenue': 796,
'commercial': 146,
'provisions': 736,
'precautions': 688,
'things': 907,
'keep': 491,
'concurrence': 160,
'admitted': 21,
'receive': 758,
'consequence': 175,
'administration': 18,
'practice': 687,
'obtained': 616,
'frequently': 394,
'right': 799,
'nomination': 601,
'entitled': 296,
'divided': 261,
'turn': 936,
'necessarily': 590,
'attend': 77,
'persons': 666,
'appears': 54,
'difficult': 242,
'principal': 705,
'addition': 14,
'formation': 382,
'privileges': 709,
'rights': 800,
'prescribe': 693,
'concerning': 157,
'respects': 792,
'capacity': 108,
'regulate': 763,
'lay': 509,
'coin': 140,
'money': 578,
'jurisdiction': 487,
'paper': 642,
'long': 535,
'declared': 222,
'land': 500,
'places': 669,
'strength': 868,
'england': 291,
'till': 917,
'propriety': 727,
'considerations': 181,
'conclusion': 158,

'confederation': 167,
'preserve': 696,
'journal': 481,
'examples': 323,
'ancient': 47,
'confederacies': 164,
'source': 854,
'institutions': 472,
'principle': 706,
'distinct': 255,
'separate': 829,
'empire': 284,
'immediate': 442,
'gradually': 412,
'independence': 457,
'sovereignty': 857,
'powerful': 685,
'tranquillity': 924,
'wars': 971,
'species': 858,
'degrees': 228,
'almost': 32,
'anarchy': 46,
'emperor': 283,
'diet': 239,
'judiciary': 486,
'tribunals': 929,
'controversies': 199,
'troops': 931,
'duties': 270,
'mutual': 582,
'value': 958,
'capacities': 107,
'exclusive': 327,
'resolutions': 787,
'apply': 58,
'safety': 807,
'territory': 903,
'support': 884,
'constitute': 186,
'europe': 310,
'natural': 586,
'fundamental': 399,
'community': 150,
'external': 351,
'dangers': 216,
'history': 431,
'oppression': 633,
'requisitions': 786,
'altogether': 38,
'side': 840,
'commonly': 149,
'half': 420,
'opposite': 631,
'parties': 648,
'defense': 225,
'pride': 704,

'views': 963,
'ready': 753,
'prejudices': 692,
'supported': 885,
'treasury': 925,
'justice': 489,
'subjects': 872,
'produced': 717,
'districts': 259,
'organization': 636,
'fully': 398,
'execute': 328,
'remedy': 773,
'city': 130,
'exercise': 332,
'soon': 853,
'fit': 372,
'possession': 682,
'asked': 73,
'entirely': 295,
'weakness': 974,
'formidable': 385,
'interest': 474,
'connected': 171,
'causes': 115,
'feeble': 365,
'whilst': 981,
'imagined': 441,
'interested': 475,
'various': 960,
'direct': 246,
'neighbors': 596,
'stability': 860,
'together': 920,
'simple': 842,
'dependent': 235,
'required': 783,
'regular': 762,
'bound': 94,
'moment': 576,
'violent': 965,
'league': 510,
'since': 843,
'left': 514,
'business': 100,
'separation': 830,
'provinces': 734,
'says': 813,
'expense': 340,
'governments': 410,
'advantage': 22,
'notwithstanding': 603,
'establishments': 309,
'efforts': 278,
'error': 302,
'depend': 233,
'ambition': 41,

'event': 312,
'besides': 87,
'suggested': 880,
'beyond': 90,
'individuals': 461,
'domestic': 263,
'popular': 676,
'bias': 91,
'expected': 338,
'particularly': 647,
'credit': 211,
'importance': 446,
'protection': 730,
'usually': 952,
'partial': 645,
'result': 795,
'due': 268,
'administered': 17,
'towards': 922,
'clearly': 137,
'lie': 525,
'favorable': 362,
'rarely': 750,
'prevail': 701,
'errors': 303,
'sacrifice': 805,
'counties': 205,
'sufficiently': 879,
'collective': 141,
'prosperity': 729,
'objects': 610,
'effect': 274,
'happiness': 425,
'proceedings': 715,
'seat': 816,
'pursued': 743,
'limits': 532,
'encroachments': 287,
'act': 7,
'zeal': 999,
'evil': 319,
'magistracy': 539,
'added': 13,
'difficulties': 243,
'serious': 831,
'willing': 985,
'alarm': 31,
'unless': 945,
'trial': 928,
'absolutely': 3,
'thirteen': 910,
'contained': 192,
'fixed': 374,
'supply': 883,
'appear': 52,
'formed': 383,
'say': 812,

'best': 88,
'thus': 916,
'subordinate': 874,
'additional': 15,
'advantages': 23,
'greatest': 416,
'dependence': 234,
'usurpation': 953,
'defects': 224,
'utility': 955,
'urged': 948,
'institution': 471,
'questions': 747,
'raised': 749,
'relative': 769,
'extent': 350,
'hold': 432,
'courts': 210,
'numbers': 605,
'good': 407,
'rest': 794,
'judicial': 485,
'certainly': 118,
'valuable': 957,
'republic': 779,
'departments': 232,
'perceive': 660,
'wealth': 975,
'active': 8,
'neither': 597,
'possible': 683,
'regarded': 761,
'indispensable': 459,
'security': 819,
'essential': 304,
'pass': 651,
'way': 972,
'duty': 271,
'delegated': 229,
'peculiar': 657,
'decisions': 220,
'discretion': 249,
'done': 264,
'impracticable': 451,
'derived': 236,
'preference': 691,
'accordingly': 5,
'whenever': 979,
'observation': 612,
'prove': 731,
'ill': 439,
'information': 465,
'reflection': 759,
'tendency': 900,
'account': 6,
'voice': 967,
'classes': 134,

'tend': 899,
'universal': 944,
'uniform': 941,
'fatal': 360,
'hazard': 427,
'inconveniences': 453,
'readily': 752,
'variety': 959,
'demand': 230,
'competent': 153,
'smaller': 849,
'come': 143,
'celebrated': 116,
'minority': 574,
'proposition': 726,
'impeachments': 445,
'arguments': 65,
'resort': 788,
'usurpations': 954,
'latitude': 505,
'evident': 317,
'theory': 904,
'applicable': 56,
'apt': 63,
'determination': 238,
'want': 969,
'faction': 354,
'equity': 301,
'wisdom': 986,
'observed': 614,
'extensive': 349,
'sensible': 827,
'intrusted': 478,
'examined': 321,
'cognizance': 139,
'district': 258,
'matters': 552,
'appeals': 51,
'appellate': 55,
'divide': 260,
'parts': 649,
'exist': 335,
'defect': 223,
'preservation': 695,
'foundation': 389,
'remain': 770,
'taxation': 896,
'regulations': 767,
'obtain': 615,
'jury': 488,
'admiralty': 19,
'review': 797,
'brought': 99,
'inquiry': 467,
'facts': 355,
'agreed': 29,
'contend': 193,

'others': 638,
'example': 322,
'provide': 732,
'prudence': 737,
'imagine': 440,
'unnecessary': 946,
'old': 624,
'world': 993,
'although': 37,
'finally': 369,
'consequently': 177,
'internal': 477,
'add': 12,
'taxes': 897,
'operations': 628,
'probably': 712,
'adequate': 16,
'effectual': 276,
'pay': 655,
'countries': 206,
'exigencies': 334,
'imposed': 449,
'energy': 290,
'project': 719,
'confederate': 166,
'structure': 870,
'satisfactory': 811,
'wise': 987,
'call': 102,
'life': 526,
'tax': 895,
'necessities': 592,
'compact': 151,
'distinction': 256,
'leave': 512,
'idea': 438,
'taking': 893,
'progress': 718,
'supplied': 882,
'opinions': 630,
'led': 513,
'known': 498,
'extreme': 352,
'obliged': 611,
'hope': 433,
'dissolution': 253,
'property': 723,
'connecticut': 172,
'exposed': 344,
'forward': 387,
'trade': 923,
'equality': 299,
'laid': 499,
'passions': 654,
'decide': 217,
'enough': 293,
'distinguished': 257,

```
'minds': 572,
'actually': 11,
'enlightened': 292,
'condition': 162,
'republics': 781,
'greece': 417,
'work': 992,
'member': 560,
'definition': 226,
'cities': 127,
'votes': 968,
'violence': 964,
'wish': 988,
'love': 537,
'fall': 358,
'weaker': 973,
'fewer': 367,
'composing': 156,
'keeping': 492,
'salutary': 809,
'critical': 213,
'imports': 448,
'exports': 343,
'rome': 801,
'distance': 254,
'alterations': 36,
'enumerated': 297,
'maxims': 553,
'criminal': 212,
'slaves': 847}
```

```
In [302... # retrain with 1000 max-features
nv = BernoulliNB()
# train Naive Bayes
nv.fit(train_X, train_labels)
```

```
Out[302... BernoulliNB()
```

```
In [303... # evaluate Naive Bayes

# with training set
print("Training accuracy", nv.score(train_X, train_labels))

# with validation set
print("Training accuracy", nv.score(val_X, val_labels))
```

```
Training accuracy 1.0
Training accuracy 0.9411764705882353
```

Logistic Regression

```
In [304... from sklearn.linear_model import LogisticRegression
```

```
In [305... class_count = texts['author'].value_counts()
class_count /= sum(class_count)
class_weight = {label : weight for label, weight in
                 zip(list(class_count.index), class_count)}
```

Base line

```
In [306... # no parameter
log_reg = LogisticRegression()
```

```
In [307... log_reg.fit(train_X, train_labels)
```

```
Out[307... LogisticRegression()
```

```
In [308... # evaluate Naive Bayes

# with training set
print("Training accuracy", log_reg.score(train_X, train_labels))

# with validation set
print("Validation accuracy", log_reg.score(val_X, val_labels))
```

```
Training accuracy 0.7121212121212122
Validation accuracy 0.5882352941176471
```

Tuning

```
In [309... log_reg = LogisticRegression(multi_class = 'multinomial',
                                C = 1,
                                class_weight = class_weight)
```

```
In [310... log_reg.fit(train_X, train_labels)
```

```
Out[310... LogisticRegression(C=1,
                             class_weight={0: 0.18072289156626506, 1: 0.1325301204819277
2,
                             2: 0.5903614457831325, 3: 0.06024096385542168
6,
                             4: 0.03614457831325301},
                             multi_class='multinomial')
```

```
In [311... # evaluate Naive Bayes

# with training set
print("Training accuracy", log_reg.score(train_X, train_labels))

# with validation set
print("Validation accuracy", log_reg.score(val_X, val_labels))
```

```
Training accuracy 0.5909090909090909
Validation accuracy 0.5882352941176471
```

Compared to Naïve Bayes, Logistic Regression does not reach the highest accuracy based on my experiments. However, my tuning for Logistic Regression succeeded in reducing overfitting.

Neural Network

```
In [312... from sklearn.neural_network import MLPClassifier
```

```
In [313... clf = MLPClassifier(hidden_layer_sizes = (200, 50),  
                      activation = 'relu', solver = 'lbfgs',  
                      random_state = 1234, verbose = True,  
                      validation_fraction = 0.1,  
                      early_stopping = True, alpha = 10e-4,  
                      tol = 10e-5, max_iter = 300)
```

```
In [314... clf.fit(train_X, train_labels)
```

```
Out[314... MLPClassifier(alpha=0.001, early_stopping=True, hidden_layer_sizes=(200, 50),  
                      max_iter=300, random_state=1234, solver='lbfgs', verbose=True)
```

```
In [315... # evaluate MLP  
  
# with training set  
print("Training accuracy", clf.score(train_X, train_labels))  
  
# with validation set  
print("Validation accuracy", clf.score(val_X, val_labels))
```

```
Training accuracy 1.0  
Validation accuracy 0.8235294117647058
```

MLP Classifier is likely to perform competently with Naïve Bayes. By adjusting solver = 'lbfgs' due to the small dataset and lowering the L2 penalty, the validation accuracy reaches at 82%.