

## Documentatie

Hieronder zijn screenshots te zien waaronder het resultaat van het programma, de uitkomst van getrainde bigram matrices, en elke prediction van elke test regel in volgorde. Het resultaat van het programma laat alleen 72 van de 72 nederlandse regels zien, dat komt omdat het bestand met test regels een lege regel had die verwijderd moest worden om het programma te laten werken. Het programma heeft een accuracy van 100%

```
92     with open('verhaal.txt', 'r') as file:
93         text = file.read().replace('\n', ' ').lower()
94     NLmatrix.train_matrix(text)
95
96     # make predictions on the test phrases
97     with open('testzinnen.txt', 'r') as file:
98         testdata = file.read().lower().splitlines()
99
100    print("Making predictions...")
101    predictor = language_predictor([ENmatrix, NLmatrix])
102    result = list(map(predictor.predict, testdata))
103
104    # results
105    print("Amount predicted as English: {}".format(result.count("ENG")))
106    print("Amount predicted as Dutch: {}".format(result.count("NL")))
107
```

letterfrequency dip (1) ×

C:\Users\User\venv\Scripts\python.exe "C:/Users/User/Documents/letterfrequency dip.py"

Training matrix: English

Training matrix: Dutch

Making predictions...

Amount predicted as English: 119

Amount predicted as Dutch: 72

Process finished with exit code 0

Run | TODO | Version Control | Terminal | Python Console

and Plugin Updates: PyCharm is ready to update. (today 01:31)

Type here to search

PC File Edit View Navigate Code Refactor Run Tools VCS Window Help GitHub [C:\Users\User\Documents\GitHub] - C:\Users\User\Documents\letterfrequency dip.py

C:\Users\User\Documents\letterfrequency dip.py

Project

- GitHub C:\Users\User\Documents\GitHub
- External Libraries
- Scratches and Consoles

letterfrequency dip.py

```
72 text = ' '.join(text.split()) # removes co
73
74
75 #fill data
76 reduce(self.__fill, text) # fills the matr
77 self.df["Total"] = self.df.sum() # adds th
78 self.df = self.df.loc[:, self.all_letters].
79 with pd.option_context('display.max_row',
None): # more optio
trainedmatrix > train_matrix() > with pd.option_context('displa
```

Run: letterfrequency dip (1) x

C:\Users\User\venv\Scripts\python.exe "C:/Users/User/Documents/letterfrequency dip.py"

Training matrix: English

|   | a        | b        | c        | d        | e        | f        | g        | \ |
|---|----------|----------|----------|----------|----------|----------|----------|---|
| a | 0.000000 | 0.029446 | 0.019242 | 0.044898 | 0.000000 | 0.007872 | 0.015452 |   |
| b | 0.051565 | 0.079190 | 0.000000 | 0.000000 | 0.270718 | 0.000000 | 0.000000 |   |
| c | 0.147059 | 0.000000 | 0.005252 | 0.000000 | 0.323529 | 0.000000 | 0.000000 |   |
| d | 0.010304 | 0.001085 | 0.000000 | 0.017354 | 0.100868 | 0.001085 | 0.011388 |   |
| e | 0.054631 | 0.002937 | 0.009790 | 0.068729 | 0.033483 | 0.005091 | 0.007441 |   |
| f | 0.059202 | 0.000000 | 0.000000 | 0.000000 | 0.068211 | 0.059202 | 0.000000 |   |
| g | 0.087866 | 0.000000 | 0.000000 | 0.002092 | 0.126569 | 0.001046 | 0.010460 |   |
| h | 0.135474 | 0.000738 | 0.000000 | 0.000000 | 0.549280 | 0.000738 | 0.000000 |   |
| i | 0.006367 | 0.003184 | 0.093385 | 0.081005 | 0.024761 | 0.014149 | 0.028652 |   |
| j | 0.067797 | 0.000000 | 0.000000 | 0.000000 | 0.203390 | 0.000000 | 0.000000 |   |
| k | 0.014184 | 0.000000 | 0.000000 | 0.000000 | 0.368794 | 0.000000 | 0.000000 |   |
| l | 0.075000 | 0.000543 | 0.000000 | 0.052174 | 0.156522 | 0.029348 | 0.000543 |   |
| m | 0.138677 | 0.027990 | 0.001272 | 0.000000 | 0.296438 | 0.003817 | 0.000000 |   |
| n | 0.014745 | 0.002647 | 0.021172 | 0.198488 | 0.077127 | 0.003025 | 0.157278 |   |
| o | 0.002286 | 0.000980 | 0.005552 | 0.013063 | 0.003266 | 0.082626 | 0.006205 |   |
| p | 0.111517 | 0.003656 | 0.000000 | 0.000000 | 0.191956 | 0.000000 | 0.001828 |   |
| q | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |   |
| r | 0.070683 | 0.001898 | 0.009488 | 0.036053 | 0.189279 | 0.000949 | 0.019924 |   |
| s | 0.086764 | 0.000404 | 0.005650 | 0.000000 | 0.116223 | 0.000000 | 0.000000 |   |
| t | 0.026551 | 0.000000 | 0.004595 | 0.000000 | 0.066888 | 0.000511 | 0.000511 |   |
| u | 0.007538 | 0.009213 | 0.058626 | 0.023451 | 0.053601 | 0.000838 | 0.041039 |   |
| v | 0.016722 | 0.000000 | 0.000000 | 0.000000 | 0.785953 | 0.000000 | 0.000000 |   |
| w | 0.252988 | 0.000000 | 0.000000 | 0.007968 | 0.118526 | 0.002988 | 0.000996 |   |
| x | 0.108696 | 0.000000 | 0.065217 | 0.000000 | 0.130435 | 0.000000 | 0.000000 |   |
| y | 0.004121 | 0.005495 | 0.001374 | 0.001374 | 0.038462 | 0.000000 | 0.000000 |   |
| z | 0.031250 | 0.000000 | 0.000000 | 0.000000 | 0.593750 | 0.000000 | 0.000000 |   |
|   | 0.136317 | 0.032146 | 0.037335 | 0.030722 | 0.013733 | 0.031231 | 0.022584 |   |

  

|   | h        | i        | j        | k        | l        | m        | n        | \ |
|---|----------|----------|----------|----------|----------|----------|----------|---|
| a | 0.003499 | 0.070262 | 0.001458 | 0.013120 | 0.114869 | 0.025364 | 0.177259 |   |
| b | 0.000000 | 0.114180 | 0.003683 | 0.000000 | 0.104972 | 0.001842 | 0.000000 |   |
| c | 0.165966 | 0.006303 | 0.000000 | 0.059874 | 0.014706 | 0.000000 | 0.000000 |   |

Run TODO Version Control Terminal Python Console

IDE and Plugin Updates: PyCharm is ready to update. (today 01:31)

Type here to search

