# Table of features

Let x be authors

Let n be emails

Let w be words

Let fw be function words

Let c be characters

Let d be digits

Let s be space

Let sen be sentence

Let p = paragraph

Let punc be punctuation

Let b number of beginnings per email

Represent total in something as E[n]

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Description | Mathematical Representation | Done |
| Lexical features | | |  |
| Average Length of email in characters | Average number of characters in emails by an author | N1[c]+ N2(c) + … + Nn[c]/n | Yes |
| Ratio of Numeric density | The ratio of digits to total no. of characters in email an author’s emails | d:c |  |
| Ratio of space characters to email length | The ratio of space character to total length | S:W |  |
| Normalized character count | Normalized frequencies of each character | 1000c |  |
| Document length in words | Average number of words per email by an author | N1(w)+N2(w)+…+Nn(w)/n | Yes |
| Average length of words | Average word length used by an author | N1[c]+ N2(c) + … + Nn[c]/N(W) | Yes |
| Average sentence length | Average size of sentence in email of an author | N1(Sen) + N2(Sen) + … + Nn(Sen) /N(n) | Yes |
| Ratio of short words (<3 characters) | Ratio of short words i.e. 3 character and less to total number of email words | N1(w) + N2(w) +…+Nn(w) / N(w) | Yes |
| Structural features | | |  |
| Number of paragraphs in documents | Average number of paragraphs in the emails of an author | N1(p) + N2 (p) + .. + Nn(p) / N(n) | Yes |
| Number of sentences in paragraph | Average number of sentences in the emails of an author | N1(sen) + N2 (sen) + .. + Nn(sen) / N(n) |  |
| Indentation of paragraphs | How paragraphs are usually indented by the author (tabs/space count) | N1(S) + N2 (S) + .. + NN(S) / N(W) |  |
| Use of greetings in beginning | Use of specific greetings in beginning of the emails of an author | B/n |  |
| Syntactic features | | |  |
| Frequencies of function words | Normalized frequencies of each function word in the emails of an authors | N(fw)/n |  |
| Frequencies of punctuations | Normalized frequencies of punctuation used by an author | N(punc)/n |  |
| Author Sentiment Analysis | The sentiment property returns a namedtuple of the form Sentiment(polarity, subjectivity). The polarity score is a float within the range [-1.0, 1.0]. The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective. | Textblob | Yes |