Training Dataset-Sentiment140 (16 million)

Positive value-8 million

Negative value-8 million

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Linear SVC | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 3112 | 2261 | 0.62 | 0.73 | 0.53 |  | |
| Negative | 1113 | 1964 | 0.33 | 0.26 | 0.46 |  | |
| Avg/total | |  | 0.54 | 0.61 | 0.51 | 0.514556 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Multinomial NB | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 3112 | 2412 | 0.65 | 0.74 | 0.57 |  | |
| Negative | 1113 | 1813 | 0.33 | 0.27 | 0.44 |  | |
| Avg/total | |  | 0.56 | 0.62 | 0.54 | 0.537514 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Logistic Regression | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 3112 | 2499 | 0.66 | 0.74 | 0.59 |  | |
| Negative | 1113 | 1726 | 0.33 | 0.27 | 0.42 |  | |
| Avg/total | |  | 0.57 | 0.62 | 0.55 | 0.54721 | |

Dataset: Twitter Global warming

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Linear SVC | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 182 | 192 | 0.76 | 0.74 | 0.79 |  | |
| Negative | 177 | 167 | 0.74 | 0.77 | 0.72 |  | |
| Avg/total | |  | 0.75 | 0.76 | 0.75 | 0.754874 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Multinomial NB | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 182 | 179 | 0.77 | 0.78 | 0.76 |  | |
| Negative | 177 | 180 | 0.77 | 0.76 | 0.77 |  | |
| Avg/total | |  | 0.77 | 0.77 | 0.77 | 0.76880 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Logistic Regression | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 182 | 189 | 0.79 | 0.78 | 0.81 |  | |
| Negative | 177 | 170 | 0.78 | 0.79 | 0.76 |  | |
| Avg/total | |  | 0.79 | 0.79 | 0.79 | 0.78551 | |

Dataset: Sentiment140 Test data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Linear SVC | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 422 | 856 | 0.36 | 0.27 | 0.54 |  | |
| Negative | 1219 | 785 | 0.59 | 0.75 | 0.49 |  | |
| Avg/total | |  | 0.53 | 0.63 | 0.50 | 0.500304 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Multinomial NB | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 422 | 765 | 0.35 | 0.27 | 0.49 |  | |
| Negative | 1219 | 876 | 0.63 | 0.75 | 0.54 |  | |
| Avg/total | |  | 0.56 | 0.63 | 0.53 | 0.5265 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Logistic Regression | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 422 | 813 | 0.35 | 0.27 | 0.51 |  | |
| Negative | 1219 | 828 | 0.61 | 0.75 | 0.51 |  | |
| Avg/total | |  | 0.54 | 0.63 | 0.51 | 0.51066 | |

Dataset: Apple Twitter Sentiment DFE

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Linear SVC | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 2978 | 2551 | 0.78 | 0.84 | 0.72 |  | |
| Negative | 571 | 998 | 0.22 | 0.17 | 0.30 |  | |
| Avg/total | |  | 0.69 | 0.74 | 0.65 | 0.65426 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Multinomial NB | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 2978 | 2683 | 0.80 | 0.84 | 0.76 |  | |
| Negative | 571 | 866 | 0.20 | 0.16 | 0.25 |  | |
| Avg/total | |  | 0.70 | 0.73 | 0.68 | 0.675119 | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of tweets in the Dataset | Logistic Regression | | | | | |
| predict tweets | F1 score | Precision | recall | | accuracy |
| Positive | 2978 | 2713 | 0.80 | 0.84 | 0.77 |  | |
| Negative | 571 | 836 | 0.21 | 0.18 | 0.26 |  | |
| Avg/total | |  | 0.71 | 0.74 | 0.69 | 0.6869540 | |

Dataset: Twitter Product Company