Probability of characters occurrences in the Alexa top 1 million.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Charater | Probability | Character | Probability | Character | Probability |
| a | 7.07 | n | 5.05 | 0 | 0.11 |
| b | 2.05 | o | 10.08 | 1 | 0.16 |
| c | 7.06 | p | 2.46 | 2 | 0.15 |
| d | 2.67 | q | 0.14 | 3 | 0.09 |
| e | 7.89 | r | 6.04 | 4 | 0.10 |
| f | 1.32 | s | 5.14 | 5 | 0.06 |
| g | 2.36 | t | 5.19 | 6 | 0.05 |
| h | 2.01 | u | 2.97 | 7 | 0.05 |
| i | 5.71 | v | 1.11 | 8 | 0.06 |
| j | 0.41 | w | 0.94 | 9 | 0.05 |
| k | 1.49 | x | 0.47 | - | 0.82 |
| l | 3.90 | y | 1.26 |  |  |
| m | 6.31 | z | 0.57 |  |  |

Key’s difficulty of typing

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Character | Difficulty | Character | Difficulty | Character | Difficulty |
| a | 3/10 | n | 4/10 | 0 | 4/10 |
| b | 4/10 | o | 2/10 | 1 | 5/10 |
| c | 6/10 | p | 3/10 | 2 | 6/10 |
| d | 3/10 | q | 3/10 | 3 | 6/10 |
| e | 2/10 | r | 3/10 | 4 | 6/10 |
| f | 3/10 | s | 5/10 | 5 | 6/10 |
| g | 4/10 | t | 3/10 | 6 | 6/10 |
| h | 5/10 | u | 2/10 | 7 | 6/10 |
| i | 2/10 | v | 5/10 | 8 | 6/10 |
| j | 4/10 | w | 3/10 | 9 | 6/10 |
| k | 5/10 | x | 7/10 | - | 6/10 |
| l | 4/10 | y | 4/10 |  |  |
| m | 3/10 | z | 6/10 |  |  |

Experiment one: Classification accuracy  
Các tiêu chí đánh giá: accuracy, precision, recall, ROC curve, true positive

rate (TPR), and false-positive rate (FPR).

A black text with black letters

Description automatically generated

A math equation with black text

Description automatically generated

A black and white math equation

Description automatically generated

A mathematical equation with black text

Description automatically generated

A graph showing the number of people in the same number

Description automatically generated with medium confidence

Đường cong ROC (Receiver Operating Characteristic) biểu thị mối quan hệ giữa tỷ lệ true positive và false positive. AUC là diện tích nằm dưới đường cong ROC và trên trục hoành (đường ngang), có giá trị trong khoảng từ 0 đến 1. Một mô hình với AUC càng gần 1 cho thấy mô hình có khả năng tốt hơn trong việc phân loại các điểm dữ liệu tích cực và tiêu cực một cách chính xác hơn, trong khi AUC gần 0.5 cho thấy mô hình phân loại không tốt hơn một cách ngẫu nhiên.

A screenshot of a computer

Description automatically generated

A graph of a function

Description automatically generated