

12. Who is the real author of Hamlet?

- Please implement a [Text Classifier](#)
- Test the [Text Classifier](#) to predict who the real author of Hamlet is.

○	○	○ Words	○ Author
○ Training	○ 1	○ W1 W2 W3 W4 W5	○ C (Christopher Marlowe)
○	○ 2	○ W1 W1 W4 W3	○ C (Christopher Marlowe)
○	○ 3	○ W1 W2 W5	○ C (Christopher Marlowe)
○	○ 4	○ W5 W6 W1 W2 W3	○ W (William Stanley)
○	○ 5	○ W4 W5 W6	○ W (William Stanley)
○	○ 6	○ W4 W6 W3	○ F (Francis Bacon)
○	○ 7	○ W2 W2 W4 W3 W5 W5	○ F (Francis Bacon)
○ Test	○ 8 (Hamlet)	○ W1 W4 W6 W5 W3	○ ?

- Please clearly shows the results of

$P(C)$ = The probability of Author C = Author C / total Author = 3/7

$P(W)$ = The probability of Author W = Author W / total Author = 2/7

$P(F)$ = The probability of Author F = Author F / total Author = 2/7

$P(W1|C)$ = The probability the Words W1 appears on the Author C document
 $= (\text{count}(W1, C) + 1) / (\text{count}(C) + \text{number of vocabulary})$
 $= (4 + 1) / (12 + 6) = 5/18$

$P(W1|W)$ = The probability the Words W1 appears on the Author W document
 $= (\text{count}(W1, W) + 1) / (\text{count}(W) + \text{number of vocabulary})$
 $= (1 + 1) / (8 + 6) = 2/14 = 1/7$

$P(W1|F)$ = The probability the Words W1 appears on the Author F document
 $= (\text{count}(W1, F) + 1) / (\text{count}(F) + \text{number of vocabulary})$
 $= (0 + 1) / (9 + 6) = 1/15$

$P(W3|C)$ = The probability the Words W3 appears on the Author C document
 $= (\text{count}(W3, C) + 1) / (\text{count}(C) + \text{number of vocabulary})$
 $= (2 + 1) / (12 + 6) = 3/18$

$P(W3|W)$ = The probability the Words W3 appears on the Author W document
 $= (\text{count}(W3, W) + 1) / (\text{count}(W) + \text{number of vocabulary})$
 $= (1 + 1) / (8 + 6) = 2/14 = 1/7$

$P(W3|F)$ = The probability the Words W3 appears on the Author F document
 $= (\text{count}(W3, F) + 1) / (\text{count}(F) + \text{number of vocabulary})$

$$= (2 + 1) / (9+6) = 3 / 15 = 1/5$$

$$\begin{aligned} P(W4|C) &= \text{The probability the Words W4 appears on the Author C document} \\ &= (\text{count}(W4, C) + 1) / (\text{count}(C) + \text{number of vocabulary}) \\ &= (2 + 1) / (12+6) = 3/18 = 1/6 \end{aligned}$$

$$\begin{aligned} P(W4|W) &= \text{The probability the Words W4 appears on the Author W document} \\ &= (\text{count}(W4, W) + 1) / (\text{count}(W) + \text{number of vocabulary}) \\ &= (1 + 1) / (8 + 6) = 2 / 14 = 1 / 7 \end{aligned}$$

$$\begin{aligned} P(W4|F) &= \text{The probability the Words W4 appears on the Author F document} \\ &= (\text{count}(W4, F) + 1) / (\text{count}(F) + \text{number of vocabulary}) \\ &= (2 + 1) / (9 + 6) = 3 / 15 = 1 / 5 \end{aligned}$$

$$\begin{aligned} P(W5|C) &= \text{The probability the Words W5 appears on the Author C document} \\ &= (\text{count}(W5, C) + 1) / (\text{count}(C) + \text{number of vocabulary}) \\ &= (2 + 1) / (12 + 6) = 3 / 18 = 1 / 6 \end{aligned}$$

$$\begin{aligned} P(W5|W) &= \text{The probability the Words W5 appears on the Author W document} \\ &= (\text{count}(W5, W) + 1) / (\text{count}(W) + \text{number of vocabulary}) \\ &= (2 + 1) / (8 + 6) = 3 / 14 \end{aligned}$$

$$\begin{aligned} P(W5|F) &= \text{The probability the Words W4 appears on the Author W document} \\ &= (\text{count}(W4, W) + 1) / (\text{count}(W) + \text{number of vocabulary}) \\ &= (1 + 1) / (8 + 6) = 2 / 14 = 1 / 7 \end{aligned}$$

$$\begin{aligned} P(W6|C) &= \text{The probability the Words W6 appears on the Author C document} \\ &= (\text{count}(W6, C) + 1) / (\text{count}(C) + \text{number of vocabulary}) \\ &= (0 + 1) / (12 + 6) = 1 / 18 \end{aligned}$$

$$\begin{aligned} P(W6|W) &= \text{The probability the Words W6 appears on the Author W document} \\ &= (\text{count}(W6, W) + 1) / (\text{count}(W) + \text{number of vocabulary}) \\ &= (2 + 1) / (8 + 6) = 3 / 14 \end{aligned}$$

$$\begin{aligned} P(W6|F) &= \text{The probability the Words W6 appears on the Author F document} \\ &= (\text{count}(W6, F) + 1) / (\text{count}(F) + \text{number of vocabulary}) \\ &= (1 + 1) / (9 + 6) = 2 / 15 \end{aligned}$$

Does d8 belong to C or W or F?

Step 1: Analysis

There are 5 Words in d8: W1, W4, W6, W5, W3

- A. The probability of d8 belonging to Author C
Applying Compare Model

$$\begin{aligned}
P(C|d8) &= P(C) * P(W1|C) * P(W4|C) * P(W6|C) * P(W5|C) * P(W3|C) \\
&= 3/7 * 5/18 * 1/6 * 1/18 * 1/6 * 3/18 \\
&= 45/1469664 = 0.000031
\end{aligned}$$

B. The probability of d8 belonging to Author W

Applying Compare Model

$$\begin{aligned}
P(W|d8) &= P(W) * P(W1|W) * P(W4|W) * P(W6|W) * P(W5|W) * P(W3|W) \\
&= 2/7 * 1/7 * 1/7 * 3/14 * 3/14 * 1/7 \\
&= 18 / 470596 = 0.000038
\end{aligned}$$

C. The probability of d8 belonging to Author F

Applying Compare Model

$$\begin{aligned}
P(F|d8) &= P(F) * P(W1|F) * P(W4|F) * P(W6|F) * P(W5|F) * P(W3|F) \\
&= 2/7 * 1/15 * 1/5 * 2/15 * 1/7 * 1/5 \\
&= 4 / 275625 = 0.000015
\end{aligned}$$

Step 2: Conclusion

Document 8 should belong to the Author W(William Stanley) , because the probability of d8 belonging to Author W(William Stanley) is the largest.