

Hamming distance to calculate
string similarity

Hamming Distance

- Number of positions in which two strings are different



Hamming Distance

- Number of positions in which two strings are different



Hamming Distance

- Number of positions in which two strings are different



Hamming Distance

- Number of positions in which two strings are different

Analytics
Vidhya

T	e	s	t
T	e	x	t

0

Hamming Distance

- Number of positions in which two strings are different

Test
Text 0

Hamming Distance

- Number of positions in which two strings are different

Test
Text 0 + 0

Hamming Distance

- Number of positions in which two strings are different

Test
Text

0 + 0



Hamming Distance

- Number of positions in which two strings are different

Test
Text

0 + 0 + 1



Hamming Distance

- Number of positions in which two strings are different

Test
Text

0 + 0 + 1

Hamming Distance

- Number of positions in which two strings are different

Test
Text

0 + 0 + 1 + 0



Hamming Distance

- Number of positions in which two strings are different

Test
Text Hamming
 distance = 1

Hamming Distance

- Number of positions in which two strings are different

Taken
Table

Hamming Distance

- Number of positions in which two strings are different

Taken
Table

Hamming
distance = 3

Hamming Distance

- Number of positions in which two strings are different

Normalized Hamming Distance

Taken
Table

Hamming Distance

- Number of positions in which two strings are different

Normalized Hamming Distance

↓

$$\frac{\text{(Hamming distance)}}{\text{len(string)}}$$

Taken
Table

Hamming Distance

- Number of positions in which two strings are different

Normalized Hamming Distance

↓

Taken
Table

(Hamming distance)

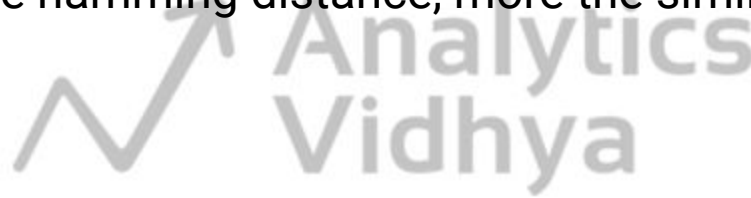
len(string)

↓

$\frac{3}{5}$

Hamming Distance

- Number of positions in which two strings are different
- Lesser the hamming distance, more the similarity



Hamming Distance

- Number of positions in which two strings are different
- Lesser the hamming distance, more the similarity
- Classical implementation handles strings of same length

Hamming Distance

- Number of positions in which two strings are different
- Lesser the hamming distance, more the similarity
- Classical implementation handles strings of same length
- Padding can be added at prefix or suffix to make strings of same length