Vladimir ID

Vladimir ID

Challenge:

1. Can the two languages be distinguished using a bag-of-words approach? Explain why.

No, because bag of words doesn’t take care of the order in the sequence

1. Can the two languages be distinguished using a bigram or trigram-based approach? Explain why.

Not always, because the incorrect could be in form like 1a1**c1234b**1d1 but in cases when the distance between c and b is 1 character, we could catch it.

1. Can the two languages be distinguished using a convolutional neural network? Explain why.

Yes, because conv net could pass to the next if there is appearance of ‘b’, ‘bc’ or ‘c’. If it passed c and after find a b it will mark it as a bad example.