# START OF QUIZ
## Student ID:
## 55343529,Nadal,Jacob

# Question 1

Topic: Lecture 8
Source: Lecture 8

What are two advantages of using .py files over .ipynb files for deployment, and two reasons why .ipynb files are preferred for prototyping or development? (1)

# Question 2

Topic: Lecture 6
Source: Lecture 6

What kinds of tags might be useful in the following text (describe at least two): "But you liked Rashomon!" "That's not how I remember it!" (1)

# Question 3

Topic: Lecture 8
Source: Lecture 8

If you were to encounter an alien text, which encoding might you want to use to digitize it? Explain briefly. (1)

# Question 4

In class, we built a POS tagger that tries to give a majority tag to a word; if it's out-of-vocabulary, it backs-off to Regexes. This is clearly overly simplistic. List two assumptions that are being violated by this model. (1)

# Question 5

Topic: Lecture 5
Source: Lecture 5

List one advantage that regular expressions have over string comparison, and one disadvantage to using them. (1)

# Question 6

Topic: Lecture 6
Source: Lecture 6

Consider using XML to represent a machine learning model's architecture. What XML tags might be useful for representing layers, activation functions, and connections between layers (you don't need to describe a deep-learning architecture - describe one you're familiar with)? If this doesn't seem possible, explain why not. (2)

# Question 7

Imagine we have a spell-checker that can identify common misspellings of words by replacing certain letters with a capture group that contains letters that are nearby on the keyboard. How aggressive of a regex would we want to write for this (ie, how many letters in the word would we want to replace with a group)? Explain. (2)

# Question 8

Do you think that we could do lemmatization before machine translation? Provide 1 argument that for why it might help, and one for why it might make things more complicated. List any assumptions that might make your answer more complicated. (2)

# Question 9

Topic: Long
Source: Lecture 8

Imagine that you find an important file buried on a hard drive found in the basement of a university. You are trying to access the data, but realize it is corrupted. Some of the bits have been flipped (switched from 0 to 1, or 1 to 0), and others have been completely deleted. You don't know the encoding, and you don't know the language the data is written in. What are some tests you could run to try to establish and restore at least some of the data? (Hint: remember that a "byte" is 8-bits, and that UTF-8 is 1 byte, or 8 bits, UTF-16 is 2 bytes, or 16 bits, and UTF-32 is 4 bytes, or 32 bits). (3)

# END OF QUIZ