

1- Which techniques you have used while cleaning the data if you have cleaned it?

-- I use some techniques

- At first, I drop duplicate and check for NAN values.
  - Use regular expression that get word has alpha letter.
  - Use punctuation in String library to remove punctuation.
  - Remove stop words using STOPWORDS library.
  - Remove any word that length less than 3 characters.
- 

2- Why have you chosen this classifier?

- I try some model like Nive Bayes, Logistic, SVM, and Neural Network and I choose Nive Bayes at the end.
  - used Multinomial Naive Bayes, because this works well for data which can easily be turned into counts, and can be run locally not like Neural Network, and train time is small time.
- 

3- How do you deal with (Imbalance learning)?

- Use the right evaluation metrics (classification report - confusion matrix).
- 

4- How can you extend the model to have better performance?

- Improve performance with data if we have more data.
- Select best algorithm can deal with data.
- Also use algorithm tuning to select best hyperparameter.

5- How do you evaluate your model?

- Use accuracy for train and test data to sure train not under or over fitting.
- F1 Score conveys the balance between the precision and the recall.
- Use Recall to measure of how many truly relevant results are returned.

-----

6- What are the limitations of your methodology or where does your approach fail?

- Because data is small, the prediction will be not good, if input job description not in train data or so far for train data.
- Also, data do not have same sample for every Job Title, so model will predict some of category well and another not good.

*By Mohamed Elsayed Mohamed Eldsoky*