Leadbook Data Challenge

Problem Description: In Leadbook we have millions of contacts and thousands of new contacts added daily. One of the important tasks is to validate and clean the job titles of contacts.In this challenge we are looking to identify the completely invalid job titles from the list of job titles.

Examples:

Valid Job Titles	**Invalid Job Titles**:
1. Data Engineer 2. Account Analyst 3. Director Network Engineering 4. Field Consultant 5. Import Export Coordinator 6. CEO 7. Lean Engineer 8. … etc	1. fdfdfdfdfdfdfdf 2. aaaa2222 3. 932516 4. asdfasdf 5. ^^ :ddddddddddddddddddddddddd 6. uki9pl;09 7. q? 8. bgvgvgbgbyhb 9. b744 10. tkw 11. c 12. fdgh n 13. = = 14. ;) 15. asfsg gfdsfg 16. ~<> ~

To Do: Write a function which takes job title as an input and returns if it is valid or not. You can use any methods from simple python scripts to NLP and Machine Learning techniques.

Given: jobtitles.csv - 7.5k Job titles with both valid and Invalid ones. (All the job titles will be in english, ignore those are not)

Evaluation: We are looking for the below metrics.

1. How did you approach the problem (you can have more than one approaches)
2. If you are using NLP or ML please describe elaborately about your approach, what

are your features, what technique/algorithm you choose, epochs,activation function and other metrics if used

3. What is your accuracy, how did you calculated and what are the future works needs to be done to improve it

Please submit your the answers,codes in **git repo**. Explore and Enjoy the task. We would like you to come up with **estimated time to complete**. Let us know if you need further information.