First I performed the import of the given libraries and dataset. The code has been run on Jupyter Notebook. Exploratory Data Analysis has been done on the given dataset i.e. Raw_Skill_Dataset.csv . I have gone through the Example_Technical _Skill.csv and found a pattern in which the skill is capitalized or separated by - or () sign whereas non-required starts with lowercase. As this is a machine learning assignment so I first tried to consider it a classification with the motive to use any classification algorithm but the dataset does not fulfill the requirement for that. Therefore, I had to go by capitalization technique i.e skills are capitalized others are not. In case of multiword phrases , all first letters have to be capital.

I defined a function to find if the first letter is capital or not. If it is capital, then 1 is appended, otherwise 0 is appended. Then this is combined with a raw dataset and the rows which contain zeros are filtered out.

In this way, we have helped Amanda to clean the dataset and extract technical(hard) skills.