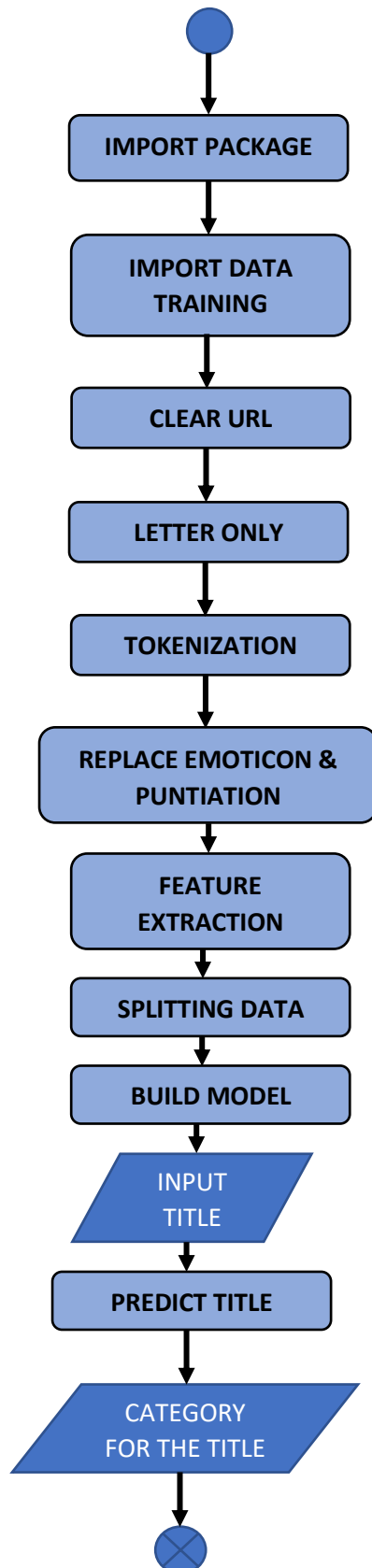


ABOUT:

Make a classifier to decide a comments whether it's spam or not. I use file "**Comment Spam.xls**" as Data Training. In the field **Class**, **0** means **not spam** and **1** means **spam**.

FLOWCHART FOR COMMENT SPAM CLASSIFICATION



NOTES

- Accuracy for the model is 86%
- This model using Convolutional Neural Network algorithm
- For feature extraction, I am using global vector (glove) to transform text to vector.
- I am using different tokenizer for this, in this task I'm using TweetTokenizer. The reason is because in this task, there are so many emoticon that people use in their comment, so I think this kind of tokenizer can help.

