

Being a detective is one of the hardest jobs in life. You don't get paid enough , you risk your life every time you are chasing or getting close to any dangerous suspect but this is nothing compared to the feeling you get that you made the city less dangerous with every criminal you put behind bars. My name is detective Cormbo Santiago. I was introduced to Natural Language Processing (NLP) by my partner who had spent sometime learning how to create and train a good model of NLP. Then we had a very good idea about how to make use of this awesome technology in order to make more criminals crack and confess on their partners. But here comes the question why do I need to use NLP ? The answer is if you reach a point to tell both suspects that the other one did confess on you and said that you did it all on your own in order to get a reduced sentence he will crack and tell you the truth , but how can you make him believe you ? you simply give him a written confession , the trick here is how can you make this written confession be written in a very close way to how the other suspect usually write stuff, the handwriting problem you can get over it by 2 means whether to give him a printed copy or by using an expert of handwritings to imperfect copying the handwriting of the other suspect . The other problem is to make what's written sound like the other suspect. Here comes the role of NLP , if you train your model good enough then it will simply output the text in a similar way of the data it was used to train the model which in this case the chats of the suspect.

The process is a four stage process, if you do it correctly you will achieve great results.

The first stage is to identify your suspects, find links between them and the motive that made them do this.

Once you reach this point you now need to start investigating them, get as many chats and written notes from them as you can. This can be achieved by simply asking for their mobile phones to get investigated , collect as much written things by them as you can to anyone , you don't need a certain or a specific chat , you need to collect as much as written data as you can.

The third stage is the complicated one , it contains 2 sub stages in order to fully be able to create a reliable NLP algorithm.

The first sub-stage is called pre-processing the data you need first to check how many languages does your suspect use , to do so you can use the langdetect library in Python. It's better if you check that all the content you have is text by using the isalpha function in python , for digits and emojis you need to replace them with text. Once you finish this process you need to clean the data you collected from any common words that is used by everyone , this common words is usually defined by stop words. For this task you can simply call the ENGLISH_STOP_WORDS from the sklearn.feature_extraction.text library in python also. Once you have removed all your stop words you can now go on with the upcoming step , start removing punctuation then you need to make all your text in lower case then you remove any brackets or emojis from the chats. Every step of them have a certain function that you need to pass your text to in python.

Once you pre-processed your data, we get to the second sub-stage which is training the model using either the `tfidf` or `CountVectorizer`. You can find both of them in the library called `sklearn.feature_extraction.text`. You will need to choose which function you are going to use then you will need to split your data in train data and test data. You can do this using `train_test_split` found in the library `sklearn.model_selection` this will split your data into test data and train data so you can go to next step. In general you will train your model using the data train data and then you will ask your model to predict some text or tell you if the text you provide it is more likely to be written by your suspect or not this process will be carried using the test data. This process is done by fitting and transforming your train data. Then you compare the predicted answers to the real outcomes. This process can be done by using Multinomial Naive Bayes classifier, you create a Multinomial Naive Bayes classifier and then fit it to your trained data then you ask it to predict the outcome on the test data then you will have to measure your accuracy score and check if your model is accurate enough or not. You can check the accuracy score using the function `metrics.accuracy_score` from the library `sklearn`. Now the closer your accuracy score to 1 the better is your algorithm. Now you can start having a conversation with your NLP and make it process the confession you want !.

We come to the final stage of our process , now you just get to the 2 suspects show them the written confession and tell them that they are going to jail for a very long time since the other already confessed and then you leave them with their thoughts and the confession. 90% of the time the suspect will crack and tell you the full truth and give you all the information you needed so you can be 100% sure of what happened and all it took you here is simply to use this nice piece of technology to get them talking by simply tricking them into thinking that the other one betrayed the other to get away or to get a reduced sentence, you make use of one of the most useful instinct of a criminal , instinct of revenge , he won't like the idea that he will take the fall all alone , so he will simply tell you the full truth.

In conclusion, NLP gave us a new way today to be able to put criminals behind bars , it gave us a new way to break the secret promises between suspects that they won't talk and this will make them always not sure if the other person talked or not so he will keep thinking should I believe and take him down with me or should I just take the fall all on my own? But who in his right mind would accept to take the fall for everything if he get betrayed ? It's totally against the human nature , which will make his instinct of revenge start controlling him and this will be the point where he is going to crack and tell you everything !.