Pipeline:-

1. **Data Cleaning:**- we first removed the hex values, alphanumerics character, stopwords from sentence column.
2. **Splitting of Data:**- We split the data into two parts one in which sentence has the label and others which don't. We give labels from 1 to 6 in the labeled dataset based on the column index (Growth - 1: Other - 6). We also add one column Index in the dataset which is useful when merging labeled and unlabelled datasets.
3. **Feature Engineering:**- we then convert the corpus of documents into vectors using tf-idf transformation.
4. **Modeling:**- We first split the labeled dataset into train and test(80:20) and run the Support Vector Classifier(SVC) on it, which gives approximately 96% accuracy. Then we train our SVC on the whole labeled dataset and predicted labels for the unlabelled dataset.
5. **Saving Back:-** we then assign a label to the original dataset and save it as Output.csv, using Index(created in step 2).

Assumption:-

We run SVC classifier because according to this paper(Multinomial Naive Bayes for Text Categorization Revisited)SVC is one of the best classifiers for such tasks.