**Text Classification Using LSTM**

**PROBLEM STAEMENT**

In this project we will be building a text classifier using LSTM and wor2vec.

**DESCRIPTION OVERVIEW**

Text classification is the task of assigning a set of predefined categories to free-text. Text classifiers can be used to organize, structure, and categorize pretty much anything. For example, new articles can be organized by topics, support tickets can be organized by urgency, chat conversations can be organized by language, brand mentions can be organized by sentiment, and so on.

There are many approaches to automatic text classification, which can be grouped into three different types of systems:

* Rule-based systems
* Machine Learning based systems
* Hybrid systems

Deep learning algorithms such as Word2vec and Glove are also used in order to obtain better vector representations for words and improve the accuracy of classifiers trained with traditional machine learning algorithms.

Few typical applications of text classification technology including all of the following:

* Social media monitoring.
* Brand monitoring.
* Customer service.
* Voice of customer.

**TECHNOLOGY USE**

Here we will be using **Anaconda Python 3.6 , Pytorch 1.4 with GPU support CUDA 10 with CuDNN 10.**

**INSTALLATION**

Installation of this project is pretty easy. Please do follow the following steps to create a virtual environment and then install the necessary packages in the following environment.

**In Pycharm it’s easy**

1. Create a new project.

2. Navigate to the directory of the project

3. Select the option to create a new new virtual environment using conda with python3.6

4. Finally create the project using used resources.

5. After the project has been created, install the necessary packages from requirements.txt file using the command pip install -r requirements.txt

**In Conda also it’s easy**

1. Create a new virtual environment using the command

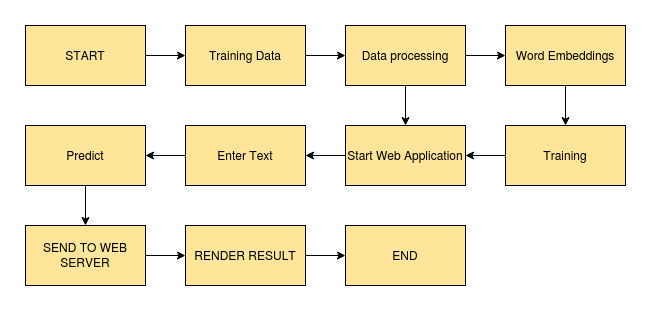
conda create -n your\_env\_name python=3.6

2. Navigate to the project directory.

3. Install the necessary packages from requirements.txt file using the command

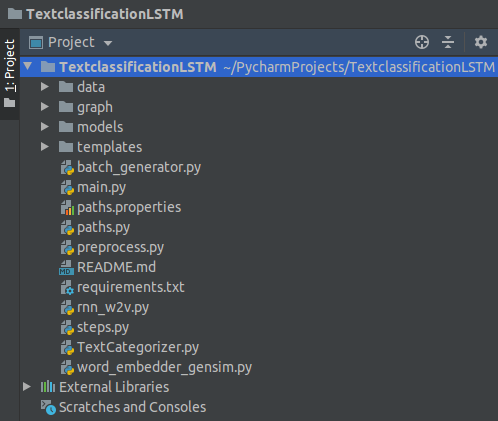
pip install -r requirements.txt

**WORKFLOW DIAGRAM**

****

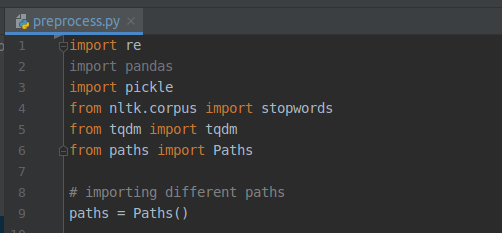
**IMPLEMENTATION**

**1. Project Directory**



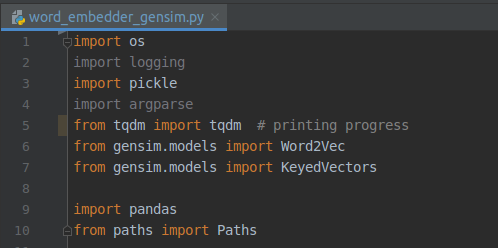
This is the complete folder stucture of the project.

**2. preprocess.py**



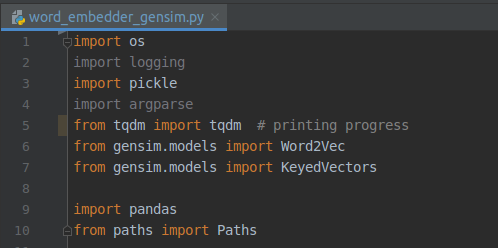
This file is used for data processing. It will create train\_preprocessed.pickle validation\_preprocessed.pickle and test\_preprocessed.pickle files under data folder.

**3. word\_embedder\_gensim.py**

****

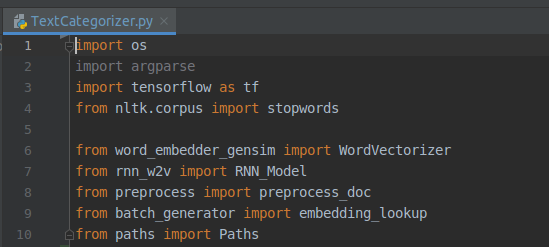
This file will training the Word2Vec embeddings.

**4. rnn\_w2v.py**

****

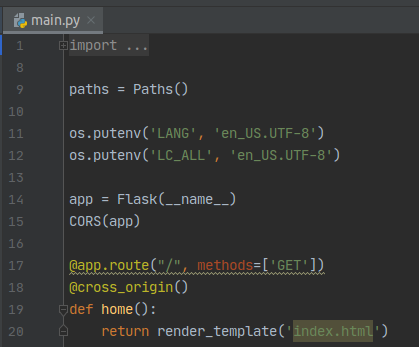
This file will train the LSTM network.

**5. TextCategorizer.py**

****

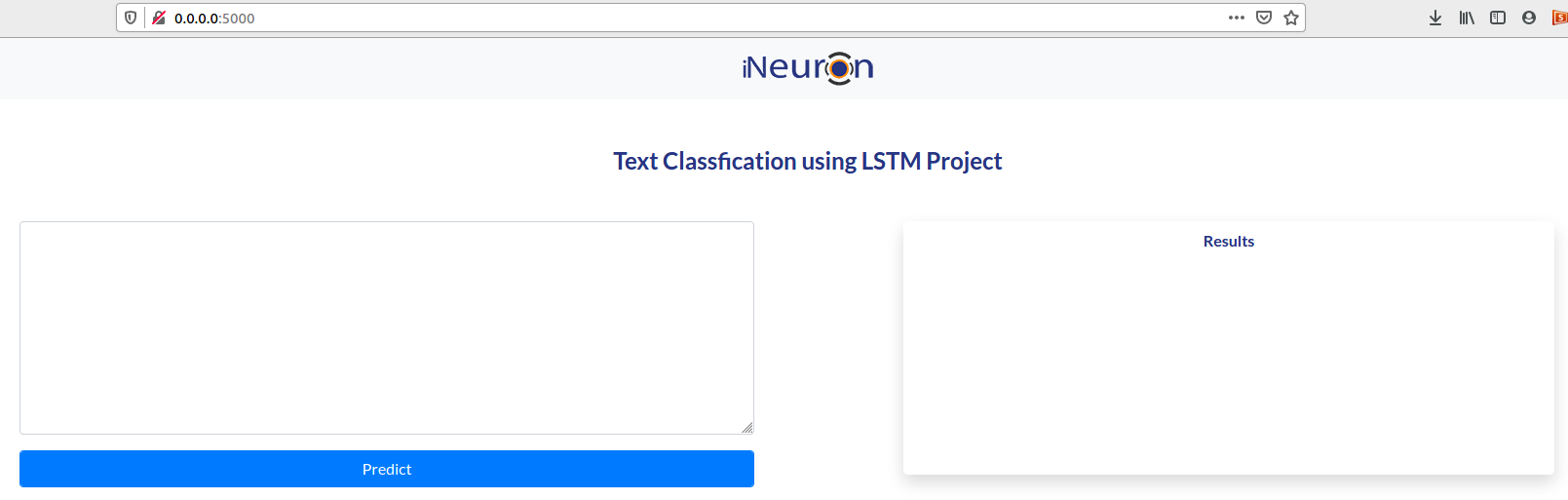
This file will be used for prediction of any input text.

**5. main.py**

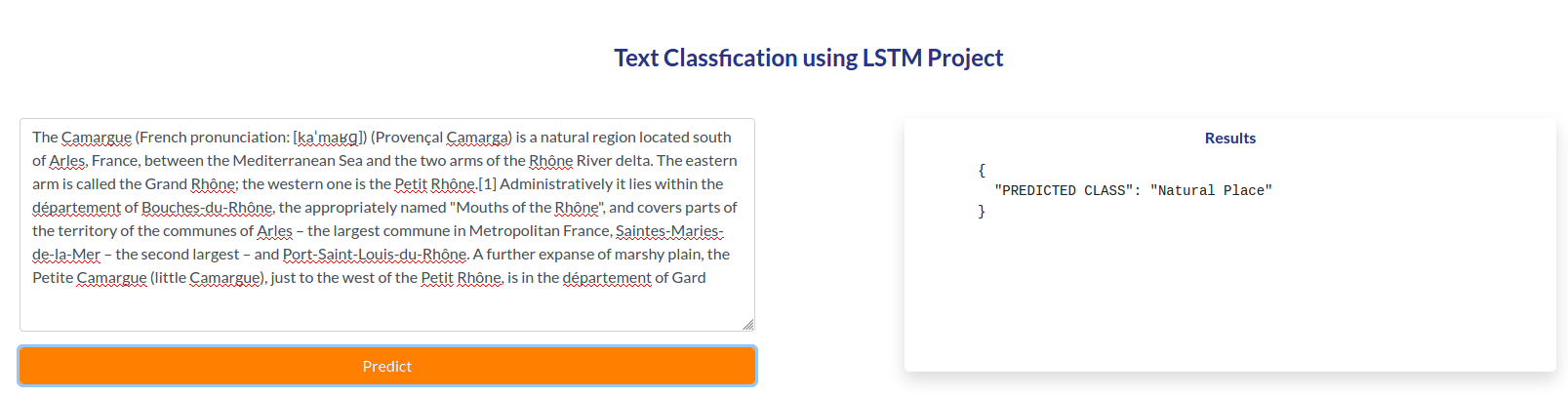


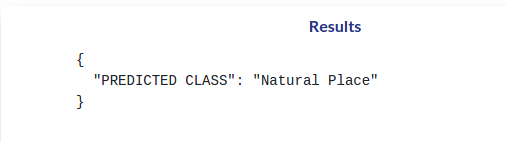
**TESTING IN LOCAL/API**

To do the test testing we need to run the main.pyand after that web server will start at <http://0.0.0.0:5000/>



Enter the text to be classified and click on Predict button.





**CONCLUSION**

Hence we have successfully the build the text classifier using Word2vec and LSTM.

**COMPARISION**

Here we have kept the scope a bit small but you can get better results using pretrained model BERT or GPT2 which are gaining a lot of popularity recently and better word embedding tecniques.