Architecture

**News Article Sorting**

Contents

1. Introduction……………………………………………………………………………….1
   1. What is Low-Level design document ?......................................1
   2. Scope……………………………………………………………………………………..1
2. Architecture………………………………………………………………………………..2
3. Architecture Description……………………………………………………………..3
   1. Data Description……………………………………………………………………...3
   2. Web Scrapping………………………………………………………………………..3
   3. Data Transformation………………………………………………………………..3
   4. Data Insertion into Database…………………………………………………….3
   5. Export Data from Database……………………………………………………….3
   6. Data pre-processing………………………………………………………………….3
   7. Data Clustering………………………………………………………………………….3
4. Model building……………………………………………………………………………….4
   1. Data from user………………………………………………………………………….4
   2. Data Validation…………………………………………………………………………4
   3. Data clustering…………………………………………………………………………..4
5. Deployment…………………………………………………………………………………….5
6. Introduction
   1. What is Low-Level design document?

The goal of LLD or a low-level design document (LLD) is to give the internal logical design of the actual program code for News article classification. LLD describes the class diagrams with the methods and relations between classes and program specs .It describes the modules so that the programmer can directly code the program from the document.

* 1. Scope

Low-Level design (LLD) is a component – level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requitement analysis and then refined during data design work

1. Architecture steps
2. start
3. data for performing tasks
4. web scraping
5. data transformation
6. data pre-processing
7. NLP Techniques
8. Data Clustering
9. Model building
10. Pushing app to cloud
11. Application start
12. Data from user
13. Data validation
14. Data clustering
15. End

3.Architecture Description

3.1 Data Description

News article dataset is the biggest publicly available dataset. It contains

Three columns namely id ,text and label

3.2 Web Scraping

In order to create a more complete data collection we will need some

More datasets.

3.3 Data Transformation

In the transformation process , we will convert our original dataset which is in JSON format to CSV format. And will merge it with the scrapped dataset

3.4 Data Pre-processing

Data pre-processing steps we could use are NULL value handling, stop words removal, punctuation removal, Tokenization, TFIDF, Imbalanced data set handling, handling columns with standard deviation zero.

3.5 Data Clustering

K-Means algorithm will be used to create clusters in the pre-processed data. The optimum number of clusters is selected by plotting the elbow plot. The idea behind clustering is to implement different algorithms to train data in different clusters. The K-Means model is trained over pre-processed data and the model is saved for further use in prediction

4 Model building

After the clusters are created , we will find the best model for each cluster. For each cluster, algorithms will be passed with the best parameters derived from grid-search. We will calculate the AUC scores for models and select the model with best score

4.1 Data for user

Here we will collect the data from user (i.e news article )

4.2 Data Validation

Here data validation will be done, given by the user

5 Deployment

We will be deploying the model to heroku