High Level Design

**News Article Sorting**

Content

Document Version control……………………………………………………………………..2

Abstarct………………………………………………………………………………………………….4

1. Introduction………………………………………………………………………………….5
   1. Why this High-Level Design Document……………………………………5
   2. Scope………………………………………………………………………………………5
   3. Definitions………………………………………………………………………………5
2. General Description……………………………………………………………………..6
   1. Product Perspective………………………………………………….……….....6
   2. Problem statement…………………………………………………………………6
   3. Proposed solution………………………………………………….………………6
   4. Technical Requirements…………………………………………….…………..6
   5. Tools Used……………………………………………………………………..………6
3. Design Details………………………………………………………………………………7
   1. Process Flow……………………………………………………………….…………7
   2. Deployment process………………………………………………………………7
   3. Event Log……………………………………………………………………………….7
   4. Error Handling………………………………………………………………………..8
   5. Performance…………………………………………………………………………..8
   6. Reusability……………………………………………………………………………..8
4. Dashboards………………………………………………………………………………….9
   1. KPIs (Key Performance Indicators)………………………………………….9

5 Conclusion…………………………………………………………………………………..10

**Abstract**

In today’s world, data is power. With News companies having terabytes of data stored in servers, everyone is in the quest to discover insights that add value to the organization. With various examples to quote in which analytics is being used to drive actions, one that stands out is news article classification.

Nowadays on the Internet there are a lot of sources that generate immense amounts of daily news. In addition, the demand for information by users has been growing continuously, so it is crucial that the news is classified to allow users to access the information of interest quickly and effectively. This way, the machine learning model for automated news classification could be used to identify topics of untracked news and/or make individual suggestions based on the user’s prior interests.

**1 Introduction**

* 1. **Why this High-Level Design Document?**

The purpose of this High-Level Design(HLD) document is to add the necessary detail to the current project description to represent a suitable model for coding .This document is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level

The HLD will:

* Present all of the design aspects and define them in detail
* Describe the user interface being implemented
* Describe the performance requirements
  1. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture , application flow and technology architecture .The HLD uses non-technical to middly -technical terms which should be understandable to the administrators of the system

* 1. Definitions

Term Description

Database : Collections of all the information monitored

IDE : Integrated Development Environment

AWS : Amazon Web Services

1. General Description
   1. Product Perspective

The News Article sorting system is a deep NLP (natural language processing) based model which help us to categorize the news articles of BBC

* 1. Problem statement

To create an AI solution for classifying the news articles of BBC and to implement the following use cases.

* To classify the articles based on there category (ex. Business)
* To cluster the same category articles under one database

2.3 Proposed solution

The solution proposed here is an news article sorting system using deep NLP

Can be implemented to perform above mentioned use cases. In first

Case it classifies the given text of article into the category of news proposed by

the BBC. In the second case it groups the similar category of articles under one

cluster and store it in a data base

* 1. Data requirements

Data requirement completely depend on our problem statement

* We need news articles of the BBC data
* We require at least 100 – 150 articles of each category with annotation
* We should clean the data
  1. Tools used
* Python 3.8
* NumPy
* Pandas
* Flask
* Scikit-learn

1. Design Details
   1. Process Flow

For classifying the different types of news articles we will use deep NLP technique. Below steps are followed for it :-

1. Collect the data from open source websites(ex. Kaggle, Uci)
2. Training and validation on dataset
3. NLP model for text classification
4. Prediction on test data
5. Take necessary actions
   1. Deployment Process
6. create the pickle file and load the model
7. create a flask app using python and jinja templating
8. create a beautiful frontend using html, CSS and bootstrap
9. create a requirements file and procfile
10. open Heroku platform and connect it to your GitHub
11. Dump your code on to GitHub and click “deploy branch”
    1. Event Log

The system should log every event so that the user will know what process is running internally.

**Initial step-by-step description:-**

1. The system identifies at what step logging is required
2. The system should be able to log each and every system flow
3. Developer can choose logging method. You can choose database logging/file logging as well.
4. System should not hang even after using so many languages
   1. Error Handling

Should errors be encountered , an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage

* 1. Performance

The news article sorting system is used for articles classification and storing the similar articles under one database to make the retravel as easy as possible

* 1. Reusability

The code written and the components used should have the ability to be reused with no problems

1. Dashboards

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the unveiled problems that if not addressed in time could cause catastrophes of unimaginable impact

* 1. KPIs (Key Performance Indicators)
* Key indicators displaying a summary of the news articles classifying system
* Time and workload reduction using this system
* To group the similar articles under one cluster
* Easy retrieval of the required category of article

1. Conclusion

The designed news article sorting system is able to predict the category of the news article we have provided to it