High Level Design (HLD)

News Article Sorting

Revision Number:
Last date of revision:

Document Version Control

DATE ISSUED	VERSION	DESCRIPTION	AUTHORS
26-06-2023	1	Initial HLD-V1.0	Ashwani Devi

Contents

Abstract

In the fast-paced digital era, staying informed about the latest news has become an essential part of our lives. However, the sheer volume and variety of news articles available can often overwhelm readers, making it challenging to find the content that aligns with their interests. This article proposes a solution to this problem by introducing category-based sorting, enabling readers to effortlessly navigate news articles based on their preferred categories such as sports, entertainment, politics, and more.

By implementing category-based sorting, readers gain several benefits. Firstly, it saves valuable time, as they can directly access the news topics they are most interested in, avoiding the need to sift through unrelated articles. Secondly, it provides a personalized news experience, allowing readers to curate their feeds based on their specific preferences and areas of interest.

The process of category-based sorting involves utilizing advanced algorithms and machine learning techniques to automatically assign articles to appropriate categories. News organizations can leverage natural language processing and text classification models to analyze article content, headlines, and metadata to determine the most suitable category for each piece of news.

Implementing category-based sorting requires collaboration between news publishers, technology companies, and readers. Publishers should ensure that their articles are accurately tagged with relevant categories, enabling the sorting algorithms to function effectively. Technology companies can develop user-friendly interfaces and intelligent recommendation systems to facilitate seamless browsing and navigation of news content. Readers, on the other hand, can provide feedback and preferences, further improving the accuracy and relevance of the category-based sorting system.

In conclusion, category-based sorting revolutionizes the way readers consume news, providing a streamlined experience that caters to their specific interests. By implementing this approach, news organizations can increase reader engagement, satisfaction, and loyalty. As the digital landscape continues to evolve, category-based sorting emerges as a valuable tool in the pursuit of enhancing news consumption in an era of information overload.

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High Level Design (HLD) Document is to add the necessary details 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 hardware and software interfaces
- Describe the performance requirements
- Include design feature and the architecture of the project
- List and describe the non-functional attribute like:

- Security
- Reliability
- Maintainability
- Portability
- Reusability
- Application compatibility
- Resource utilization
- Serviceability

(HLD)

1.2 Scope

The HLD document presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

1.3 Definitions

NAS - News Article Sorting

2 General Description

2.1 Product Perspective

News Article Sorting system is an innovative news aggregation platform that aims to enhance the news consumption experience by providing users with a customizable and curated news feed based on their preferred categories. It offers a seamless solution for readers overwhelmed by the abundance of news articles by intelligently sorting and categorizing them into relevant topics, including sports, entertainment, politics, technology, and more.

2.2 Problem Statement

The abundance of news articles and the lack of efficient categorization options make it challenging for readers to find and consume relevant news content based on their interests, leading to information overload and decreased user satisfaction.

2.3 Proposed Solution

News Article Sorting system offers a category-based sorting system that intelligently categorizes news articles and provides a personalized news feed, enabling users to efficiently consume relevant content aligned with their interests and preferences.

2.4 Further Improvements

In order to further enhance NewsAggregator+, the following improvements can be implemented: integrating seamless social sharing capabilities, introducing smart notifications for personalized updates, supporting multimedia content for a more engaging experience, implementing language localization for broader accessibility, enabling cross-platform synchronization for a seamless user experience, incorporating data insights and analytics for informed decision-making, and integrating accessibility features to ensure inclusivity for all users.

2.5 Data Requirements

Data requirements completely depend on our problem statement.

We need data for some articles with their categories for training our model. After that we just need news articles, they can be in list form and comma-seperated and our model will be predicting the categories of those articles.

- News article: We need news articles for doing the predictions
- Article Category: Only required for training data.

2.6 Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Plotly, Flask etc are used to build the whole model.

- Visual Studio Code is also used as IDE.
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.

- Render is used for deployment of the model.
- Python, Flask is used for backend development.

2.7 Constraints

A primary constraint for the development of News Article Sorting system is the availability of limited development resources and budget. Allocating resources for extensive feature development, testing, and maintenance can be challenging, requiring careful prioritization and efficient resource utilization. Additionally, technical limitations and compatibility issues across different platforms and devices may present constraints on the implementation and seamless integration of desired functionalities. Moreover, compliance with data privacy regulations and ensuring robust security measures within the platform can impose additional constraints that need to be addressed to safeguard user data.

2.8 Assumptions

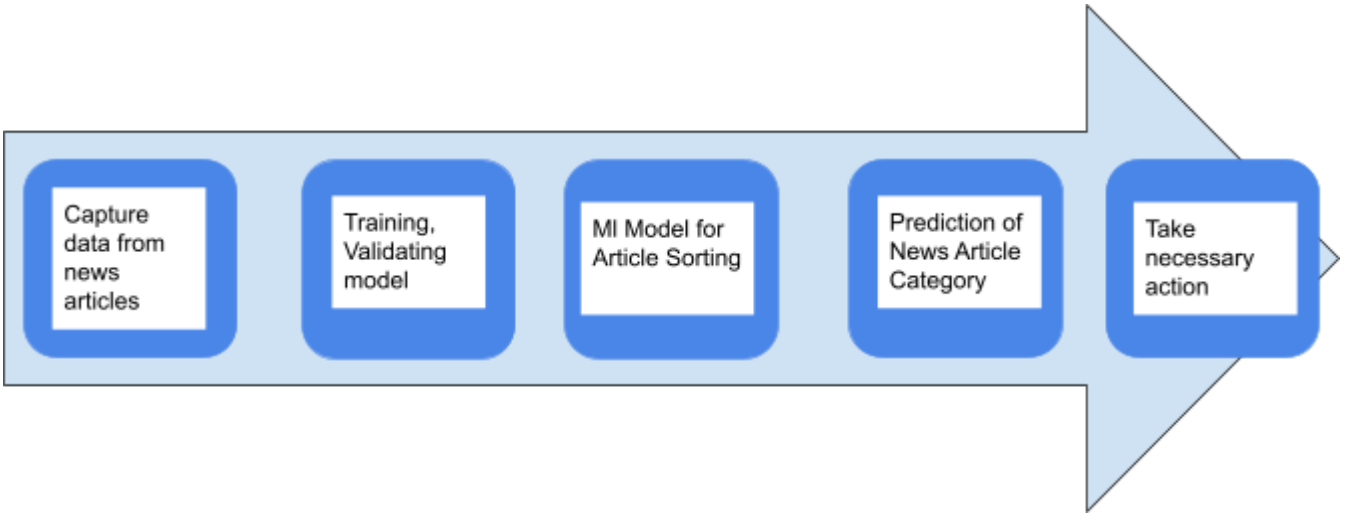
The assumption is that users have a desire for a more efficient and personalized news consumption experience and are open to utilizing a platform like News Article Sorting system to fulfill their needs.

3 Design Details

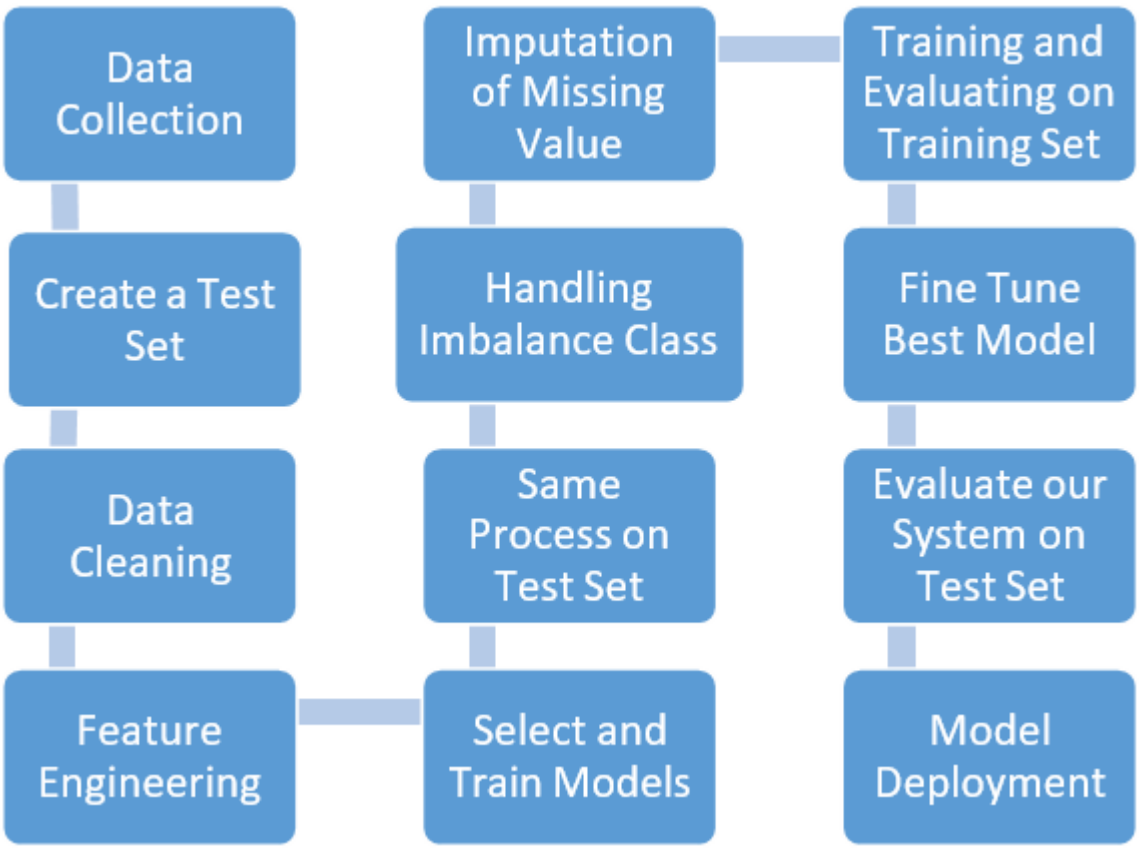
3.1 Process Flow

For predicting the News category, we will use a machine learning base model. Below is the process flow diagram is as shown below

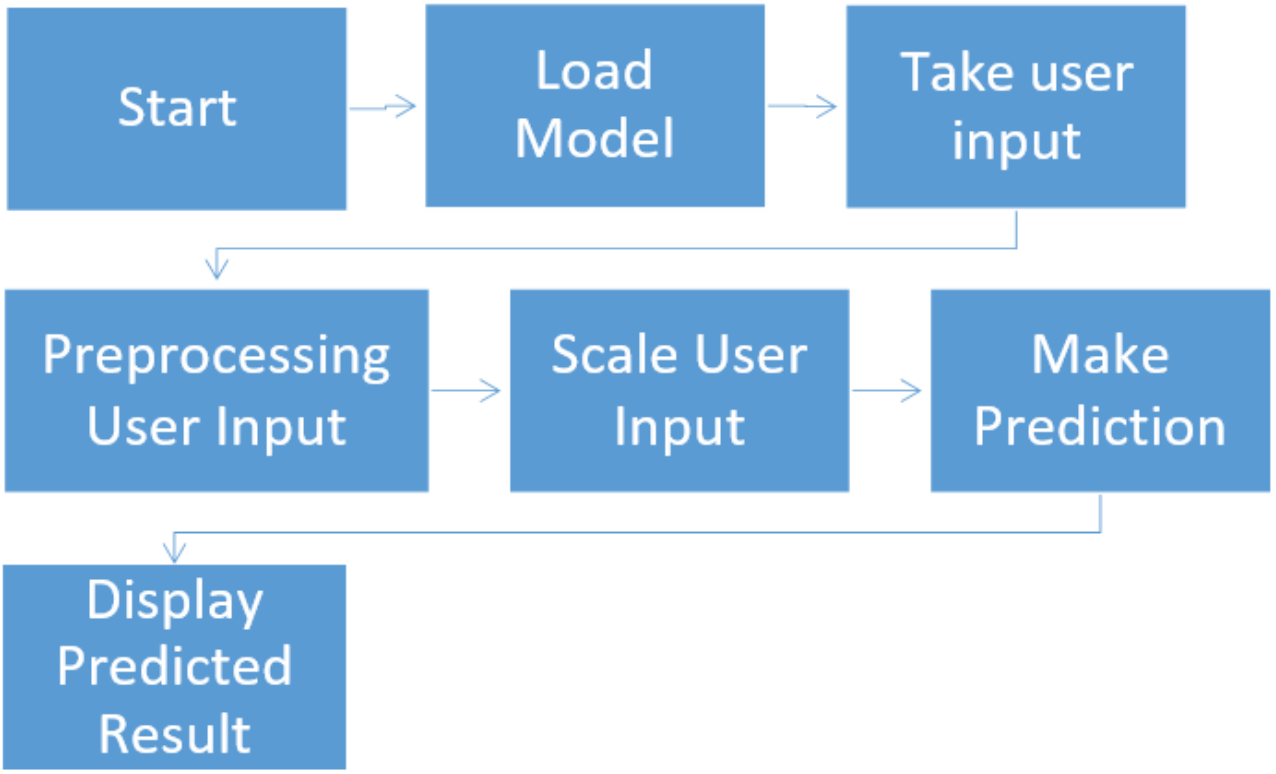
Proposed methodology



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process


Start
Load Model
Take user input
Preprocessing User Input
Scale User Input
Make Prediction
Display Predicted Result

2.6 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. Developers can choose logging methods. You can choose database logging/ File logging as well.
8. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

2.7 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.

4 Performance

The performance of the News Article Sorting System is expected to provide efficient categorization and sorting of news articles, deliver personalized recommendations, ensure seamless navigation, and offer real-time updates. The platform aims to enhance user satisfaction, engagement, and the overall news consumption experience.

4.1 Reusability

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

4.2 Application Compatibility

The different components for this project will be using python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

4.3 Resource utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

render

5 Conclusion

In conclusion, the introduction of category-based sorting in the News Article Sorting system revolutionizes the way readers consume news by offering a personalized and streamlined experience. The proposed solution provides efficient categorization, personalized recommendations, and a user-friendly interface. Further improvements such as social sharing, smart notifications, and multimedia support can enhance user engagement and satisfaction. However, constraints such as limited resources, technical challenges, and data privacy concerns need to be carefully addressed during the development and implementation process. With continued enhancements and a focus on user needs, News Article Sorting system has the potential to reshape news consumption in the digital age.

6 References

BBC News Classification Data Set

URL: https://www.kaggle.com/c/learn-ai-bbc/data