

NEWS ARTICLE SORTING

High Level Design Document



Pauline IC

ineuron.ai

Revision Number: 2.0

Last Date of Revision: Jan 16, 2023

DOCUMENT VERSION CONTROL

| Date Issued | Version | Description | Author |
|-------------|---------|--------------|------------|
| 12-01-2023 | 1.0 | HLD Document | Pauline IC |
| 16-01-2023 | 2.0 | HLD Document | Pauline IC |
| | | | |

Table of Contents

| | |
|---|----------|
| Abstract: | 3 |
| 1. Introduction | 3 |
| 1.1 <i>Why this High-Level Design Document?</i> | 3 |
| 1.2 <i>Scope</i> | 3 |
| 1.3 <i>Definitions:</i> | 4 |
| 2. General Description | 4 |
| 2.1 <i>Product Perspective</i> | 4 |
| 2.2 <i>Problem Statement</i> | 4 |
| 2.3 <i>Proposed Solution:</i> | 4 |
| 2.4 <i>Further Improvements:</i> | 4 |
| 2.5 <i>Data Requirements:</i> | 4 |
| 2.6 <i>Tools used:</i> | 4 |
| 3. Design Details | 5 |
| 3.1 <i>Process Flow</i> | 5 |
| 3.2 <i>Deployment Process</i> | 5 |
| 4. Performance | 5 |
| 4.1 <i>Reusability:</i> | 5 |
| 4.2 <i>Resource Utilization</i> | 5 |
| 5. Deployment | 6 |
| 6. Conclusions | 6 |

Abstract:

Data is power in today's world. Since news organizations have terabytes of data kept on servers, everyone is trying to find insights that will benefit the company. The classification of news articles stands out among the many examples I could give of how analytics is being utilized to motivate actions. There are many sites that produce a huge amount of news every day today on the Internet. Additionally, user demand for information has been steadily increasing, thus it is critical that the news be classified to enable users to quickly and effectively obtain the information of interest. In this project, the machine learning model for automatic news classification is discussed. This technique could be used to find untracked newspaper articles and/or provide personalized recommendations based on the user's past preferences.

1. Introduction

1.1 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 before 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 features and the architecture of the project
- List and describe the non-functional attributes like:
 - Security
 - Reliability
 - Maintainability
 - Portability
 - Reusability
 - Application compatibility
 - Resource utilization
 - Serviceability

1.2 Scope

The HLD documentation outlines the system's architecture, including the technology architecture, application architecture (layers), application flow, and database architecture. The HLD employs simple to moderately complex terms that system administrators should be able to understand.

1.3 Definitions:

| Term | Description |
|-------|-----------------------------------|
| ML | Machine Learning |
| API | Application Programming Interface |
| Flask | Framework for deploying the model |
| AWS | Amazon Web Service |
| EC2 | Elastic Compute Cloud |
| NLP | Natural Language Processing |
| SSH | Secure Shell |

2. General Description

2.1 Product Perspective

The News Article Sorting Model is an NLP model helps us to classify News Articles to various categories such as Sports, Politics, Finance, Entertainment etc.

2.2 Problem Statement

The developed NLP-based model can assist the user in receiving news article recommendations based on their prior search history. The recommendations may be in the areas of politics, finance, entertainment, sports, technology, etc.

2.3 Proposed Solution:

The proposed solution uses the BBC News article dataset. The model is trained to classify any new news article. The model has been deployed in AWS EC2 instance.

2.4 Further Improvements:

A recommendation system can be integrated with this model to make it better. The user can be benefitted by the personalized recommendations.

2.5 Data Requirements:

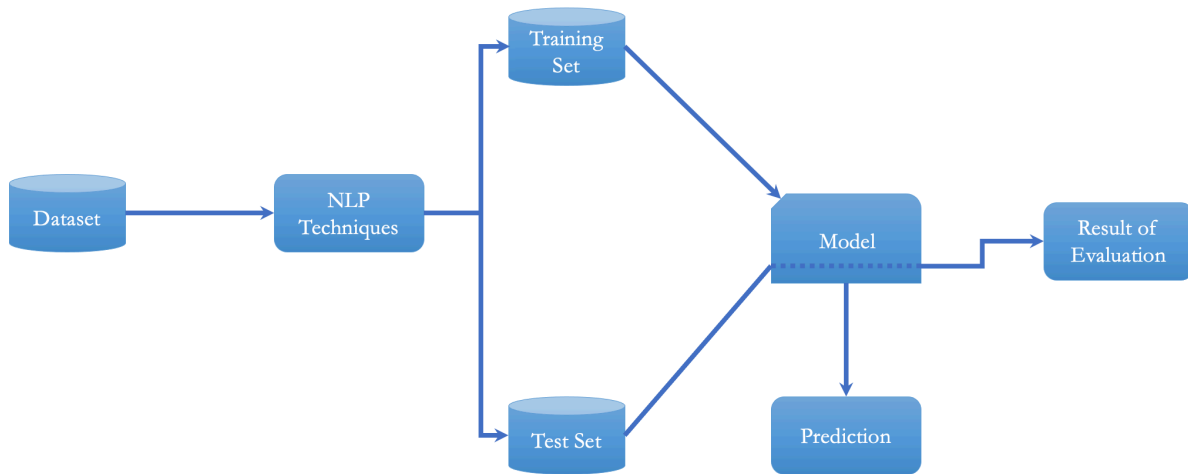
- News Articles
- Classification Group
- Database

2.6 Tools used:

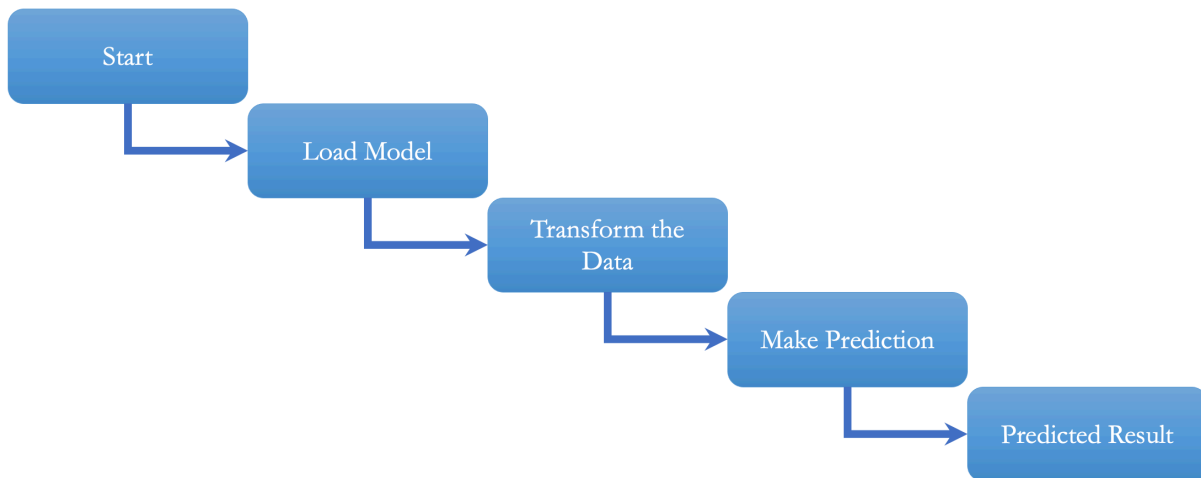


3. Design Details

3.1 Process Flow



3.2 Deployment Process



4. Performance

MultinomialNaiveBayes Algorithm predicted the results with 97% accuracy.

4.1 Reusability:

The code written and the components used have the reusing ability without any problems.

4.2 Resource Utilization

A performing task uses all the processing power available for its smooth performance.

5. Deployment



6. Conclusions

News Article Sorting will classify every news article to different categories. Based on the model's learning, this is carried out. To improve prediction, the model is trained using thousands of news articles and their classification. Any news article could be classified by the model with 97% accuracy.