

Project Title	Document Tagging
Technologies	NLP and Machine Learning
Domain	Finance
Project Difficulties level	Advance

Problem Statement:

Several thousand blog entries are actively shared on social media every hour; for example, blogging services like Tumblr¹ had over 70 billion postings across various communities by January 2014. Keywords or "#tags" (hashtags) have been shown to serve as group identities and brand labels in addition to being topic markers. Authors on Tumblr can create their own tags or use pre-existing tags to label their blogs. It can be difficult to create or choose tags for optimum dissemination, and authors may not be able to attach all of the essential tags. In recent years, algorithm-driven document tagging has emerged as a possible answer to this challenge. Blog search, clustering related blogs, displaying subjects connected with trending tags, and personalization of blog postings are just a few of the downstream applications of automatically labelling these blogs. The customization algorithm could match user interests with tags linked with a blog post to improve user engagement.

The main objective here is -

1. Tagging news articles or blog posts with relevant tags from a collection of predefined ones is coined as document tagging.
2. To provide an accurate tagging of articles which can be beneficial for several downstream applications such as recommendation and search.
3. To demonstrate the effectiveness of the approach, do experiments on several datasets and show promising results against state-of-the-art methods.

Dataset:

You have to collect dataset for this project (e.g., Wiki10, WikiNER and etc.), and based on that, you have to design your solution and create a repo for the dataset.

Project Evaluation metrics:**Code:**

- You are supposed to write a code in a modular fashion
- Safe: It can be used without causing harm.
- Testable: It can be tested at the code level.
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on GitHub.
- You have to keep your GitHub repo public so that anyone can check your code.
- Proper readme file you have to maintain for any project development.
- You should include basic workflow and execution of the entire project in the readme file on GitHub
- Follow the coding standards: <https://www.python.org/dev/peps/pep-0008/>

Database:

- You are supposed to use a given dataset for this project which is a Cassandra database.
- <https://astra.dev/ineuron>

Cloud:

- You can use any cloud platform for this entire solution hosting like AWS, Azure or GCP

API Details or User Interface:

- You have to expose your complete solution as an API or try to create a user interface for your model testing. Anything will be fine for us.

Logging:

- Logging is a must for every action performed by your code use the python logging library for this.

Ops Pipeline:

- If possible, you can try to use AI ops pipeline for project delivery Ex. DVC, MLflow , Sagemaker , Azure machine learning studio, Jenkins, Circle CI, Azure DevOps , TFX, Travis CI

Deployment:

- You can host your model in the cloud platform, edge devices, or maybe local, but with a proper justification of your system design.

Solutions Design:

- You have to submit complete solution design strategies in HLD and LLD document

System Architecture:

- You have to submit a system architecture design in your wireframe document and architecture document.

Latency for model response:

- You have to measure the response time of your model for a particular input of a dataset.

Optimization of solutions:

- Try to optimize your solution on code level, architecture level and mention all of these things in your final submission.
- Mention your test cases for your project.

Submission requirements:

High-level Document:

You have to create a high-level document design for your project. You can reference the HLD form below the link.

Sample link:

[HLD Document Link](#)

Low-level document:

You have to create a Low-level document design for your project; you can refer to the LLD from the below link.

Sample link

[LLD Document Link](#)

Architecture: You have to create an Architecture document design for your project; you can refer to the Architecture from the below link.

Sample link

[Architecture sample link](#)

Wireframe: You have to create a Wireframe document design for your project; refer to the Wireframe from the below link.

Demo link

[Wireframe Document Link](#)

Project code:

You have to submit your code GitHub repo in your dashboard when the final submission of your project.

Demo link

[Project code sample link :](#)

Detail project report:

You have to create a detailed project report and submit that document as per the given sample.

Demo link

[DPR sample link](#)

Project demo video:

You have to record a project demo video for at least 5 Minutes and submit that link as per the given demo.

Demo link

[Project sample link :](#)

The project LinkedIn a post:

You have to post your project detail on LinkedIn and submit that post link in your dashboard in your respective field.

Demo link

[Linkedin post sample link :](#)



iNeuron