

Architecture document design

(Social Media Community Using Optimized Clustering Algorithm)

Document Version Control

Date Issued	Description	Version	Author
12th Feb 2024	First Draft	1.1	Hemnath MR
13th Feb 2024	Added Workflow chart	1.2	Hemnath MR
15th Feb 2024	Added Exception Scenarios Overall,Constraints	1.3	Hemnath MR
16th Feb 2024	Added KPIs	1.4	Hemnath MR
17th Feb 2024	Added user I/O flowchart	1.5	Hemnath MR
19th Feb 2024	Added EHR, LSTM model diagrams	1.6	Hemnath MR
22th Feb 2024	Added dataset overview and updated user I/O flowchart.	1.7	Hemnath MR
23th Feb 2024	Restructure and reformat LLD	1.8	Hemnath MR

Contents

Document Version Control	2
Abstract	4
1 Introduction	5
1.1 Why this Low-Level Design Document?	5
1.2 Scope	6
1.3 Constraints	6
1.4 1.3 Risks	6
1.5 1.4 Out of Scope	7
2 Technical specifications	7
2.1 Predicting Disease	7
2.2 Logging	7
2.3 Database	7
3 Technology stack	8
4 Proposed Solution	9
5 Model training/validation workflow	10
6 User I/O workflow	12
7 Exceptional scenarios	13

Abstract :

With the rapid expansion of social media platforms and the increasing complexity of online communities, managing and analyzing user-generated content is crucial. This project explores optimized clustering algorithms to navigate social media communities. It categorizes users, posts, and interactions based on shared interests, behaviors, and demographics, enabling personalized recommendations and targeted advertising. The algorithm also incorporates real-time processing to adapt to evolving trends, fostering positive engagement, and ensuring a safer online environment.

Introduction :

1.1 Why this Architecture Design Document?

The purpose of this Architecture Design 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.

Scope :

The HLD documentation 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.

2.8.Constraints

Constraints for "Social Media Community Using Optimized Clustering Algorithm" include data privacy compliance, scalability for large datasets, algorithm performance, interpretability of results, resource optimization, data quality management, ethical considerations, and documentation for long-term maintenance.

1.4 Risks

Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

Developing a full-fledged social media platform interface is out of scope for the project.

2 Technical specifications

2.1 Dataset

Dataset	Finalized	Source
Instagram	yes	https://www.kaggle.com/

Instagram dataset overview :

The social media cluster algorithm dataset offers a comprehensive overview of engagement metrics and post metadata, enabling analysis across platforms. It provides insights into user engagement patterns and content popularity, aiding in identifying trends and influential posts.

Engagement Metrics

	Impressions	Saves	Comments	Shares	Likes	profilevisit	Follows
▶	3920	98	9	5	162	35	2
	5394	194	7	14	224	48	10
	4021	41	11	1	131	62	12
	4528	172	10	7	213	23	8
	2518	96	5	4	123	8	0
	3884	74	7	10	144	9	2
	2621	22	5	1	76	26	0

Content Categorization

	Caption	Hashtags
▶	Here are some of the most important data visua...	#finance #money #business #investing #inves...
	Here are some of the best data science project ...	#healthcare #health #covid #data #datascien...
	Learn how to train a machine learning model an...	#data #datascience #dataanalysis #dataanaly...
	Here's how you can write a Python program to ...	#python #pythonprogramming #pythonproject...
	Plotting annotations while visualizing your data i...	#datavisualization #datascience #data #dataa...
	Here are some of the most important soft skills t...	#data #datascience #dataanalysis #dataanaly...
	Learn how to analyze a candlestick chart as a d...	#stockmarket #investing #stocks #trading #m...

Audience Insights

	fromhome	fromhashtags	fromexplore	fromother
▶	2586	1028	619	56
	2727	1838	1174	78
	2085	1188	0	533
	2700	621	932	73
	1704	255	279	37
	2046	1214	329	43
	1543	599	333	25

Content Performance

	Caption	Likes	Comments	Shares
▶	Here are some of the most important data visua...	162	9	5
	Here are some of the best data science project ...	224	7	14
	Learn how to train a machine learning model an...	131	11	1
	Here's how you can write a Python program to ...	213	10	7
	Plotting annotations while visualizing your data i...	123	5	4
	Here are some of the most important soft skills t...	144	7	10
	Learn how to analyze a candlestick chart as a d...	76	5	1

4.4.Deployment

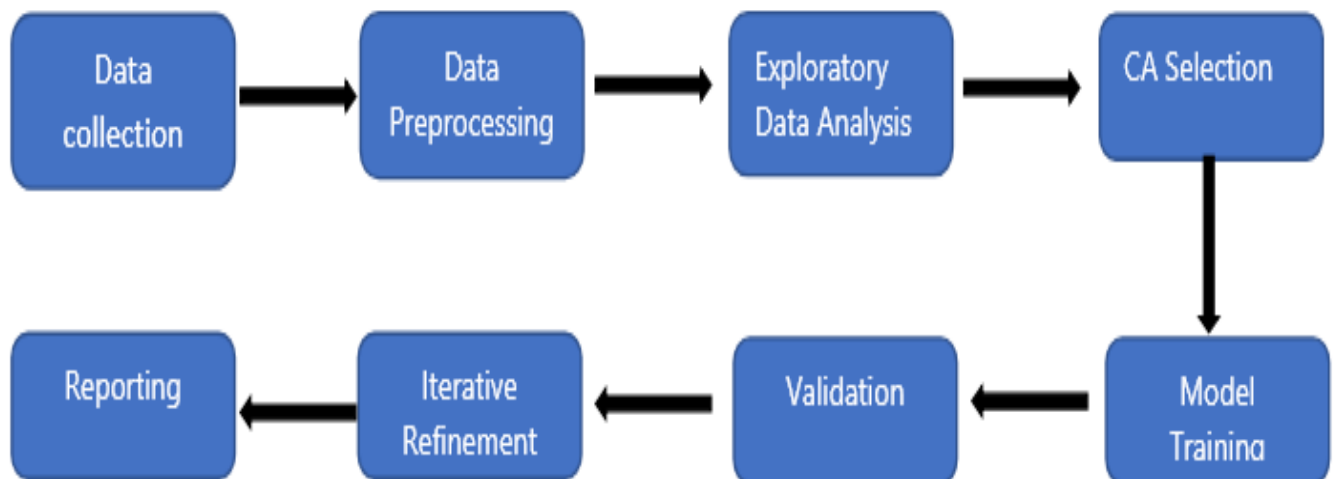


2.3.Proposed solution

The proposed solution involves collecting and preprocessing social media data, selecting clustering algorithms, and performing feature engineering. Models are trained and evaluated using metrics like silhouette score. Real-time processing enables adaptation to trends, while intuitive visualizations aid interpretation. The solution is integrated into existing platforms for seamless operation and enhanced insights into online communities.

3.1.Process flow

Here's a simplified process flow chart for the project "Social Media Community Using Optimized Clustering Algorithm"



5.1.KPIs (Key Performance Indicator)

- **Engagement Rate:** Measure the level of interaction your social media cluster algorithm generates with users. This includes likes, shares, comments, and other forms of engagement relative to the total reach or impressions of your content.
- **Audience Growth:** Track the increase in followers or subscribers across your social media platforms as a result of implementing the cluster algorithm. This metric indicates the effectiveness of your algorithm in attracting and retaining new audience members.
- **Content Relevance:** Assess the relevance of the content suggested by your algorithm to the interests and preferences of your target audience. This can be measured by analyzing feedback and time spent on content.