

Authors: Piyawat Ounruen, Apichai Srikawinnad, Kawin Tepawong and Chunakorn Danintha  
 Advisor: Dr. Kamonphop Srisopha  
 Data Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

## ABSTRACT

This study explores energy consumption challenges, focusing on battery-related issues faced by developers on Stack Overflow. Using a public Stack Overflow dataset hosted on Google BigQuery, the research employs K-means clustering algorithms, together with statistical analysis and visualization techniques, to extract meaningful insights. The study aims to identify the types of battery-related questions that are commonly discussed, those that are easy to answer, and those that remain unresolved due to a lack of accepted solutions. By analyzing these trends, the research seeks to provide a deeper understanding of the challenges developers encounter in this topic.

## INTRODUCTION

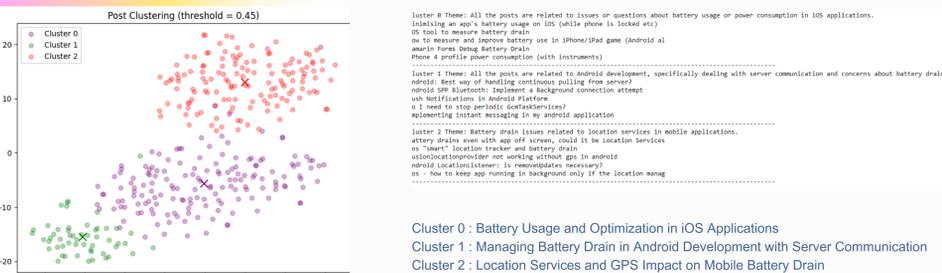
Stack Overflow is a widely used platform where software developers seek solutions to programming challenges, with battery-related issues gaining significant attention, particularly in mobile application development, embedded systems, and energy optimization. As battery-powered devices become increasingly prevalent, developers face challenges in managing power consumption, optimizing battery performance, and writing energy-efficient code. Common concerns include extending battery life, managing background processes, and utilizing battery monitoring APIs. This study analyzes battery-related discussions on Stack Overflow to identify recurring themes and key challenges. By examining these questions, the research aims to provide insights into battery optimization trends and best practices for energy-efficient software development.

## OBJECTIVES

1. To identify hidden trends and insights within the Battery-Related Discussions dataset.
2. To create data visualizations that facilitate analysis and interpretation.
3. To summarize common issues and the factors influencing their occurrences.

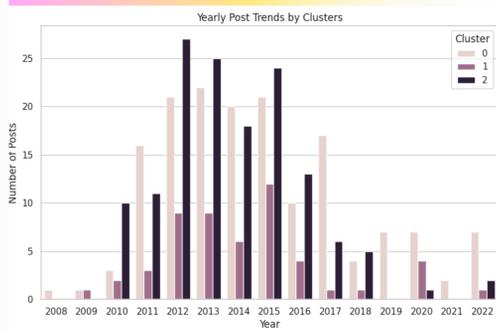
## RESULT

### Post Clustering



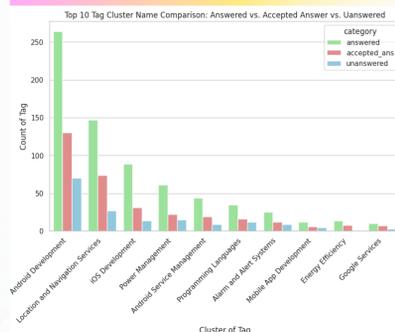
Use k-means clustering to cluster the posts and OpenAI to decide the cluster name, then replace them with more meaningful ones.

### Yearly Post Trends by Cluster



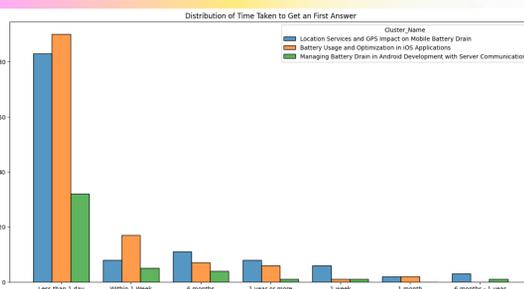
Plot the number of posts created in each year by cluster.

### Top 10 Tag Cluster Name



Plot the top 10 tags that got the most answers, accepted answers, and unanswered posts.

### Days to get first answer



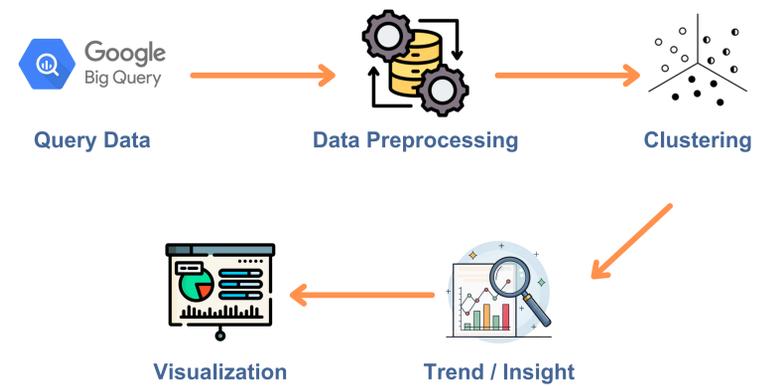
Plot the time it took to get the first answer by cluster and divide the gaps into : Less than 1 day, within 1 week, 1 Month, 6 Months, 6 Months - 1 Year, 1 Year or more. Then show the percentage of posts that got an answer within each date range.

Cluster	date_ans_diff_range	answer_count	total_answer_count	percent_first_answer_data
0	Less than 1 day	157	202	77.72277
	Within 1 week	25	202	12.37623
	6 months	0	202	0.00000
	1 year or more	0	202	0.00000
	1 week	1	202	0.00500
	1 month	2	202	0.00990
1	Less than 1 day	0	202	0.00000
	Within 1 week	40	81	49.25125
	6 months	4	81	4.93827
	1 year or more	20	81	24.69136
	1 week	1	81	1.23457
	1 month	0	81	0.00000
2	Less than 1 day	121	189	64.02116
	Within 1 week	18	189	9.52381
	6 months	10	189	5.30159
	1 year or more	12	189	6.34921
	1 week	11	189	5.82011
	1 month	7	189	3.70370
6 months - 1 year	4	189	2.11640	

## TOOLS



## PROCESS



## CONCLUSION

The analysis of the dataset using K-means clustering, based on the similarity of post topics and content, revealed that setting  $k = 3$  effectively partitions the data, clearly distinguishing between groups with differing content. This approach avoids creating clusters with overly similar content. Furthermore, clustering was not only used for post categorization but also for grouping tags with similar meanings or within the same fields, given the large number of overlapping tags (Ouni et al., 2023). Upon analyzing the three clusters, it was found that posts without an Accepted Answer accounted for 42%, particularly in topics related to Battery and Drain. This suggests that the lack of solutions that effectively resolve the issues, or the absence of an Accepted Answer selection by the post author, may contribute to this high percentage. This limitation is acknowledged as a potential constraint of the study. The graph titled Yearly Post Trends by Cluster indicates that Cluster 0 had the highest number of posts compared to Cluster 1 and Cluster 2, suggesting an increasing interest or prevalence of issues related to this topic. Furthermore, a decrease in the number of posts in 2020 was observed, likely due to advancements in AI technologies such as Google Search, Stack Overflow Auto-Suggest, and GitHub, which enable users to find answers more efficiently without needing to post new questions. The enhancement of existing databases with more comprehensive content has allowed users to retrieve solutions from previous posts, leading to a reduction in the need for new inquiries after 2015. Additionally, the analysis reveals that tags with responses within one day and those with the most responses are often related to Android Development, particularly in the areas of Location Services and Navigation Services, indicating a trend toward Green Software Development aimed at improving energy efficiency in devices, thereby reducing power consumption and enhancing battery optimization. These findings provide a deeper understanding of the challenges developers encounter in this topic.

## REFERENCES

- [1] OUNI, ALI, ET AL. "AN EMPIRICAL STUDY ON CONTINUOUS INTEGRATION TRENDS, TOPICS AND CHALLENGES IN STACK OVERFLOW." PROCEEDINGS OF THE 27TH INTERNATIONAL CONFERENCE ON EVALUATION AND ASSESSMENT IN SOFTWARE ENGINEERING. 2023.
- [2] OPENAI. "CLUSTERING." OPENAI COOKBOOK, 2024. AVAILABLE: [HTTPS://COOKBOOK.OPENAI.COM/EXAMPLES/CLUSTERING](https://cookbook.openai.com/examples/clustering).