Proposal (Group 8): An Analysis on Decentralized Social Media Platforms

Problem:

With the boom in social media platforms in recent years, most of these platforms are owned by a single parent corporation (e.g. Facebook, Instagram, WhatsApp are all owned by Facebook Inc). These corporations generate revenue from selling user data and through ads, which have led to issues concerning data privacy and data breaches.

To combat this, there has been a rise in decentralized social media platforms (community driven open source platforms not owned by a single entity). These platforms give users control of their data, and some platforms maintain revenue through crowdfunding rather than ads. However, the largest obstacle that exists with these platforms is the user base for these platforms are fairly small, and most are unheard of.

In this project, we will explore what the different communities in these decentralized platforms have to offer and what can be done to promote their growth.

Datasets:

- **Mastodon Dataset-** This dataset contains all information about Mastodon, a decentralized social media platform. The link contains data about demographics, usage over time, topics associated with each instance and more.
- **Data breaches in Centralized Social Media over time-** This dataset will be useful for elaborating on some of the weaknesses of centralized platforms and covers data breaches. (data breaches & world's biggest data breaches/hacks)
- **Diaspora Dataset:** This dataset has information about the user Demographics, usage, and topics associated with each node and more of Diaspora, a decentralized social media platform popular in 2011. (https://the-federation.info/diaspora)

Proposed Solution and Real-World Application:

The proposed solution in this project will focus on the exploration of what overlap exists in decentralized platforms with their centralized counterparts.
Questions we plan to explore that will assist us in our solution include:

1. Where are the decentralized communities distributed around the world?
2. Are there correlations between funding over time with number of users over time for centralized and decentralized platforms?
3. Data breaches over time for each social media platform?
4. What tags are used to describe the communities in decentralized platforms?

We plan to analyze collected data on these decentralized platforms through possible clustering techniques to identify communities that exist on both decentralized and centralized platforms and determine how much of an intersection there is between the two. From this

intersection, we can identify certain demographics of users on centralized platforms that can smoothly migrate over to decentralized ones, and find similar communities to the ones that they came from.

Real-world applications of this solution include recommender systems that can give users of centralized platforms insights on what decentralized communities they would best fit in, given their preferences.

Project steps

Step	Estimated completion time	Person(s) in charge
1. Extracting, merging and cleaning up data (getting the datasets ready)	Two weeks	Mohammed, Suhrid, Lucas, Richa
2. Data Visualization	One week	Lucas, Richa
3. Conclusions and Interpretations	One week	Mohammed, Suhrid