



Meeting Minutes - 31 Jan 2022



Meeting Minutes for 31 January 2022

Date:	 31 Jan 2022
Time:	 2:00 PM
Location:	Zoom
Present:	@ Ow Ling Jia @ Chen Jian Yu @ Joshua Wong
Absent:	@ ONG JHIN YEE _ @ Tian Le Cheow @ Sarah Chin

Goals

Discuss app deployment plan and data collection with sponsor.

Agenda

No.	Agenda Item	Remarks
1.	Integration of server and GPU	Jian Yu
2.	Facebook data restrictions	Ling Jia

Discussion topics

Agenda	Action By
Integration of server and GPU	

- Will incur high costs for AWS Lambda function because the processes will take very long to process
- Using EC2 will be very time consuming because we would have to learn how to configure the server and compromise our time spent on the actual analysis
- Current progress: Writing the ETL pipelines and uploading the data on the cloud database
 - Connect to remote GPU and run on our local machine, and only using AWS as a storage for the data first
 - Deploy the CPU server when Prof is back and the server is back up
- Uncertain on whether the server will be back up in time
 - Back up plan is just to deliver the modules on local machine
- How to connect CPU and GPU?
 - E.g. save pickle trained model on CPU server
 - If GPU needed, can API call GPU server directly?
 - GPU calculate (data modelling and processing) and then store everything on database
 - Only works if calling is not in real time
- Final plan: CPU and GPU scripts stored separately
 - Parts which require GPU can be stored on GPU server (e.g. crawler and modelling)
 - Do not need root access to run code on GPU server (python and other packages are installed already)
 - Raw data extract information required Build/retrain models on GPU Output new labels, etc store to NoSQL database CPU only calls for data from the database using APIs
 - App/frontend codes stored on CPU server

@ Chen Jian Yu

Facebook Data Restrictions


- For Government pages, data cannot be scraped using Selenium and we are faced with permission issues using FacePager
 - FacePager requires the account to have an authentication token which will require our account to be a verified business account
 - Need to have a business
 - Need Privacy Policy
 - Too time consuming and complicated a process
- Update: Don't worry about data that cannot be collected, and focus on what can be scraped
 - try and look for more pages and see whether we can add on to the current list
- Impact: Less data collected, but not within our control because of the restrictions imposed by the social media platform
 - Mitigation plan:
 - Cover as many of those pages from other platforms
 - Regularly check Facebook to look out for new potential sources of data and add them to our list of pages for data collection


@ Sarah Chin

:note: Meeting Notes - Updates

Update	Risk Level	Mitigation Plan
-	-	-

 Decisions

 Administrative Matters

Date of Next Meeting:	04 Feb 2022
Time of Next Meeting:	 5:00PM