Milestone 1 – Proposal week 5

Submit a proposal with the following topics covered as a PDF:

Topic – Scraped dataset from fashion websites and set topics for advertising and keywork searches.

Business Problem – The goal is the scrap data from fashion blogs, fashion sites, cosmetic sites etc. so gather their info into plain text doc and then run that through a model for topics. Once a topic is assigned to that site or page, we can target ads to that page then.

Datasets – I’m scraping the data myself. I can just put in a website and grab all text from it and store it in a text file. I haven’t really done this before, so hoping it goes well.

Methods – Its mostly going to be web scraping with automation and topic modeling. Will start with LDA and maybe move to BERT or some other kind of transformer.

Ethical Considerations – The ethical concerns will be are the topics accurate and is the client getting the value they believe. If the topics are accurate, but applied to site that doesn’t align with the company values or goals is also an issue. Also with automation, some companies don’t want their website getting scraped.

Challenges/Issues – I’ve done some web scraping, but nothing like this. I have done topic modeling, so hoping for the best. I also haven’t really used transformers much, so hoping for that best on that as well.

References – I’ll probably just look it up and see what the results are on a few of them and go from there. I’m not really sure how to check these otherwise.