CS 410 FA24: Project Proposal

Ad Personalization with Sentiment Analysis

Junyoung Lee (jl298, Coordinator)

Project Overview

The BI team at an SNS company aims to build a data pipeline for personalized ad delivery. The system utilizes widely-used, proven engines for NLP and sentiment analysis to analyze text data created by SNS users to understand their interests, select and deliver appropriate advertisements based on these insights, and establish a feedback loop process based on user 'reactions' and 'responses'.

System flow

1. Data Collection (Chained job)

- Gather user-text data such as user posts, comments, and messages from the internal database of the SNS platform

2. Text Preprocessing (Chained job)

- Refine user-text such as removing stop words, tokenization, stemming, lemmatization, and cleaning unstructured data (special characters, URLs, emojis) (Tool: NLTK or spaCy)

3. NLP (Chained job)

- Vectorize user-text as sentence-embedding (Tool: Transformers)
- Apply sentiment analysis: (1) classify sentiments and (2) provide an intensity scores (Tool: Transformers)

4. Ad Matching (Chained job)

- Apply similarity calculation w/ prebuilt ad-embeddings(vectorized ads) and weighting (Method: Cosine Similarity or dot product)
- Select ads and add them to the SNS feed, based on: (1) User-behavior-history, (2) Sentiment scores, (3) Randomization (for variety)

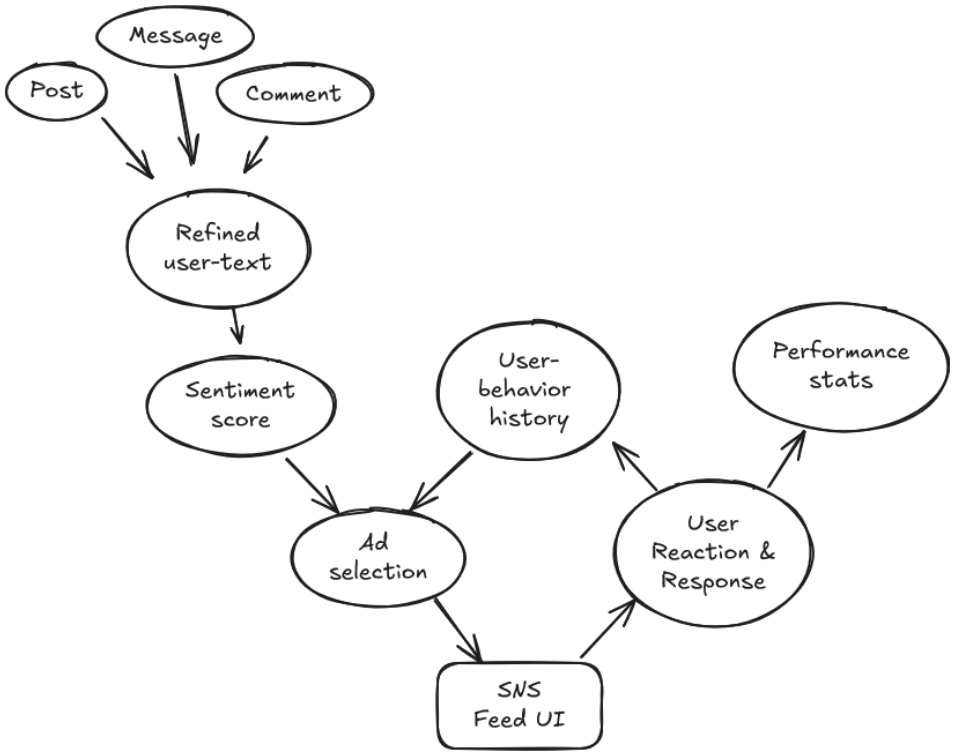
*** SNS feed (via dummy UI or emulated interactions)**

- Display ads, capture user reactions and responses, add them to the User-behavior-history

*** Performance Analysis (Separate job)**

- Measure reactions(click, etc.) to evaluate ad performance: Click-through rate, Post-click behavior, No-Reaction rate → Indirect evaluation of model
- Measure responses(Report-Ad, Ad-Survey, etc.) to evaluate sentiment analysis: Accuracy, F-score → Direct evaluation of model

Data flow



Outcome

1. [Data] Base: **User-text(Post,Comments,Messages), User-behavior-history**
2. [Data] Base: **Ad Content Data (+ Ad creatives)**
3. [Data] Result: **Matched Ad Data for SNS Feed**
4. [Data] Result: **Ad performance stats, Sentiment analysis evaluation stats**
5. [Code] Data Pipeline: **Text preprocessor, Sentiment analyzer, Ad matcher, Performance Analyzer**
6. [Code] Web UI: **Performance Analysis dashboard**
7. [Code] Web UI: **Dummy SNS feed**