Kick-off Notes

The Trade Desk.

Jiefei Ma & Chris Hawkes.

Tracking Brand Sentiment via Public News.

Ads allow users to use the internet for free

↳ Opposite is a paywall.

Ads work using Open Real-time Bidding.

↳ It's a kind of auction for ads.

SSP → Server hosting the ad. "Server side Platform".
DSP → Companies that bid. "Demand-side Platform".

Timeout after 100ms.

~ 30 million auctions per second.

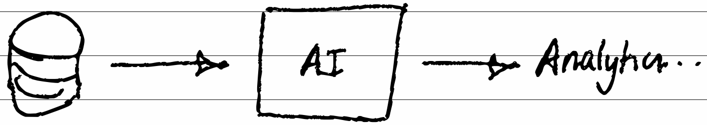
DMP → "Domain Management Platform"
Keeps track of the user's demographics
which is given to the SSP/DSPs.

Can't violate user privacy information UPI
PII → Personally Identifiable Information.

Often companies need to advertise to change public sentiment after a blunder.

Can use NLP, etc. to develop an algorithm for tracking brand sentiment

Pipeline:



Can be as simple as a scale, could be multi label?

Challenges:

- How to extract news article content from web page?
- How to identify articles mentioning a certain brand?
- How to detect sentiment in a brand?
- How to do all of them at scale? → Main ML challenge.

Available dataset → could use an API or Scraper
↳ use common crawl ⇒ commoncrawl.org.
monthly web scraping since 2014.
monthly releases 360TB of data.

Provides raw HTML and text processed formats → (latter isn't perfect).

What is Spark?

NLP libraries → John Snow Labs & Hugging Face.
Containerization → Kubeflow and Docker
ML flow → Platform for model development / registry.
ML libraries → PyTorch & Tensorflow.

Use GitHub.

MVP:

- Cloud-based data processing and ML pipelines
- Model Evaluation results
- Database for times series → S3? Mongo? Or Relational.
- Interactive Dashboard for tracking.

Stretch goals:

- Build an app that uses this data (can be a prototype).

Can start with top 500 companies (but should be general).

Stretch goal: To give reasoning why the sentiment.
Also can measure the sentiment within sentences and rank them.

Bare minimum for dashboard is selected company and time range

Can add more though

Publication is possible.

Start brain storming during holidays. Meetings initially ad hoc and then regular. Definitely meet before Week 2 of Spring.

Deadlines

4th Feb : Report 1

? : Demonstration.

2nd May : Final Report & Presentation.