EXCLUSIVE!
NEWS!

ETHICS
Engineers

Through the scope of the article "*MIND: A Large-scale Dataset for News Recommendation*", how do engagement patterns and article characteristics influence behavior in users with no interaction history versus users with history, and how effectively do LightFM and LSTUR function as equitable recommendation systems?

EXCLUSIVE!
PHOTOS!

PHOTOS!

EXCLUSIVE!

BACKGROUND

How do we get our news?

- Often, our “choices” are aided by a news recommendation system
- These recommender systems have broad ethical implications

The MIND Dataset

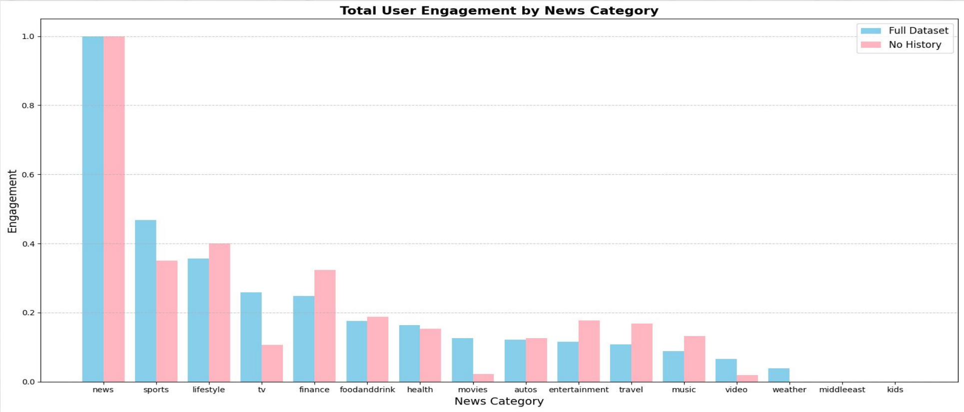
- Two datasets collected from the Microsoft News website
  - User behavior dataset – tracks the articles in a user’s history and the articles suggested to the user
  - Article dataset – describes each article’s information

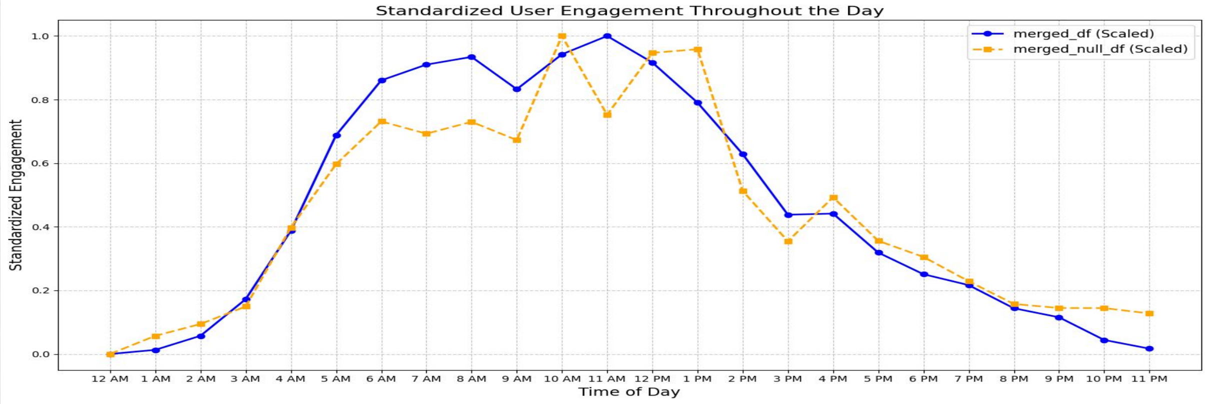
Our questions and goals

- Users with no history – how do we suggest articles to them equitably?
- Which kinds of sentiments and themes do users interact with the most?
- How do different models perform?

NEWS!

EXPLORATORY/SENTIMENT ANALYSIS





Full dataset

Sentiment	CTR
Negative	5.37
Neutral	3.68
Positive	3.27

No history dataset

Sentiment	CTR
Negative	5.27
Neutral	3.72
Positive	3.18

LIGHTFM: Recommendation Model

NEWS!

EXCLUSIVE!

Python Library for building recommendation systems that uses user and item information to make better recommendations

- Alternative to LIBFM

Full dataset	AUC	MRR	nDCG@5	nDCG@10
Category+SubCategory	92.7	11.1	4.2	4.2
Category+SubCategory+Title	71.6	1.2	0.2	0.2

Loss Type	Train AUC	Test AUC
Warp	96.4	92.7
Warp-kos	92.1	88.6
Bpr	83.6	68.4

Optimal Parameters (n_components = 30, epochs = 30) using category+subcategory item_features

No History dataset	AUC	MRR	nDCG@5	nDCG@10
Category+SubCategory	80.9	9.5	3.2	3.1
Category+SubCategory+Title	94.2	25.7	11.1	13.6

– For consistency, training values were excluded, but it’s worth noting that for all metrics the training values were consistently higher than the test values –

NEWS!

EXCLUSIVE!

LSTUR: Recommendation Model

Captures long-term and short term interests using news and user encoders

- Analyzes word interactions in news titles and relationships between different browsed articles

Full dataset	**Loss**	**AUC**	**MRR**	**nDCG@5**	**nDCG@10**
Category+SubCategory	67.9	57.7	21.5	15.4	19.9
Category+SubCategory+Title	68.1	59.6	20.6	14.4	19.6

No History dataset	**Loss**	**AUC**	**MRR**	**nDCG@5**	**nDCG@10**
Category+SubCategory	69.1	54.1	16.8	10.7	15.4
Category+SubCategory+Title	70.1	54.2	20.0	12.4	16.2

Conclusions

- Models using many predictors were accurate for users without history; models using few predictors were accurate for users with history
- An effective news recommender must account for how different these user classes are, possibly by combining multiple models

PITFALLS/LIMITATIONS

Pitfalls

- Adapting - Understanding the dataset
- LightFM – Preventing overfitting by monitoring performance on both training and test datasets
- Using proper evaluation metrics
  - attempted to keep as similar to MIND research paper as possible

Limitations

- Differences between our research and MIND research paper
  - LightFM vs LIBFM
  - train vs test dataset
  - Can’t be generalized to longer periods of time
- Computational limitations
  - Reduce data size
  - Parameter testing