NEWS RECOMMENDATIONS WITH ARTICLES AND USER LOGS

Team 27

Qiyao Wu

Jiashun Wang

Rongxiang Zhang

Rasya Soeroso

Motivation

Recommendation System

Ads

Confirmation Bias

The tendency of people to favor information that confirms their existing beliefs.

Personalized News Feed

Amplifies the confirmation bias.

Increased Traffic

More clicks, more time spent.

Make More Money

Motivation

Confirmation Bias

The tendency of people to favor information that confirms their existing beliefs.

Personalized News Feed

Amplifies the confirmation bias.

Increased Traffic

More clicks, more time spent.

Make More Money

Motivation

Confirmation Bias

The tendency of people to favor information that confirms their existing beliefs.

Personalized News Feed

Amplifies the confirmation bias.

Increased Traffic

More clicks, more time spent.

Make More Money

Motivation

Confirmation Bias

The tendency of people to favor information that confirms their existing beliefs.

Personalized News Feed

Amplifies the confirmation bias.

Increased Traffic

More clicks, more time spent.

Make More Money

Objectives

Analyzing news and user behavior data

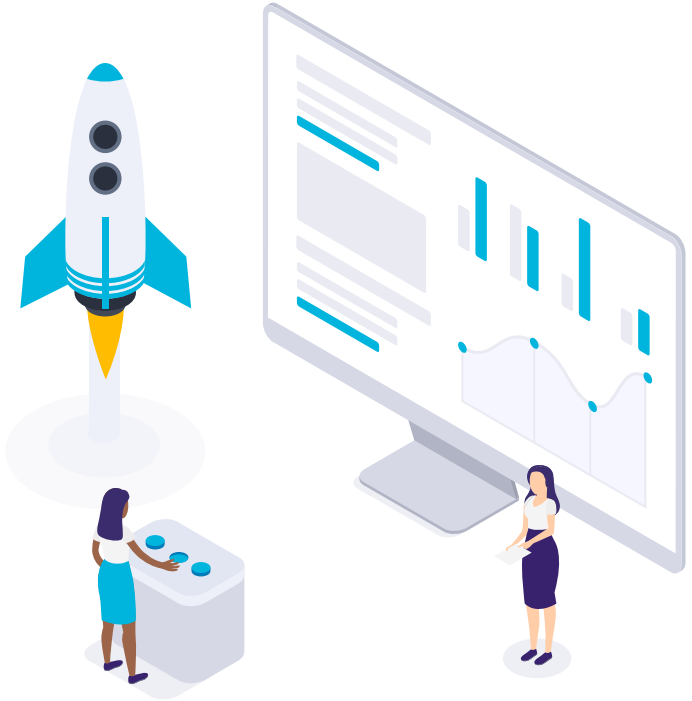
How can an ad server (e.g. Google Adsense) get more users to see ads?

Recommendation system

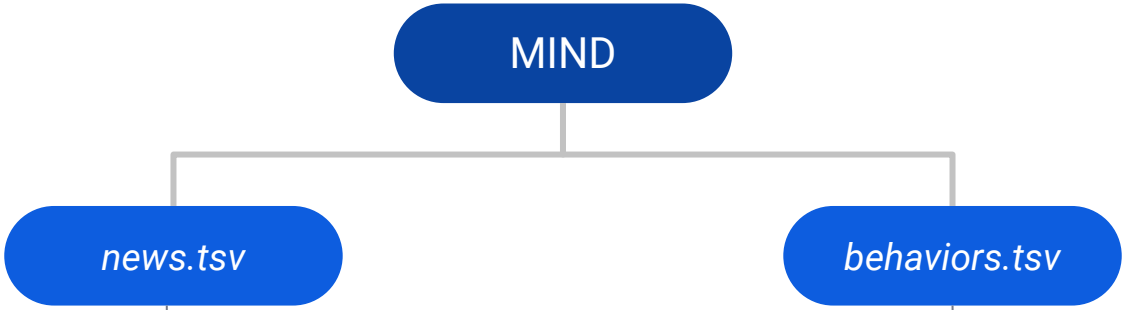
Help users to select the news they actually want to see. The more time they spend on reading more news, the more profits the platform gains.

Data Processing and Analysis

Let’s see what the data tells us!



MIND: Microsoft News Recommendation Dataset Wu, Fangzhao, et al. (2020, July) [1]



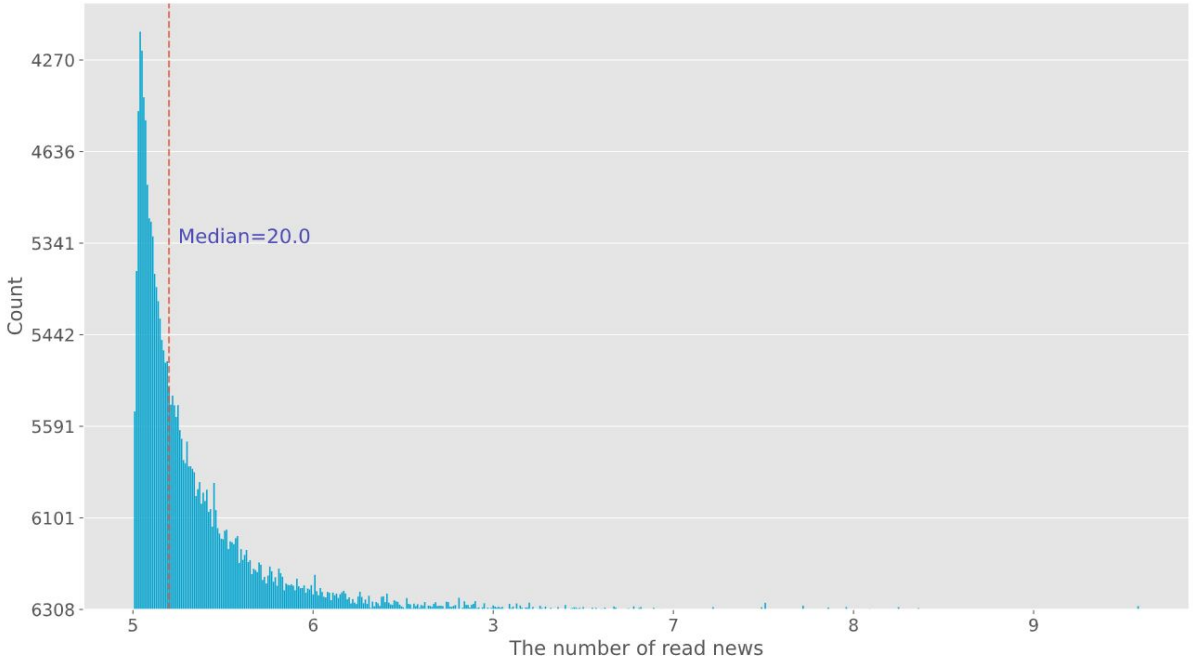
	News id	Category	SubCategory	Title	Abstract
0	N55528	lifestyle	lifestyleroyals	The Brands Queen Elizabeth, Prince Charles, an...	Shop the notebooks, jackets, and more that the...
1	N19639	health	weightloss	50 Worst Habits For Belly Fat	These seemingly harmless habits are holding yo...

	Impression id	User id	Time	History	Impressions
0	1	U13740	11/11/2019 9:05:58 AM	N55189 N42782 N34694 N45794 N18445 N63302 N104...	N55689-1 N35729-0
1	2	U91836	11/12/2019 6:11:30 PM	N31739 N6072 N63045 N23979 N35656 N43353 N8129...	N20678-0 N39317-0 N58114-0 N20495-0 N42977-0 N...

Preprocessing

- Data cleaning (i.e. remove NaN, merge duplicate user)
- Data analysis
  - News data
  - Behaviors data
  - Make connections between news data and user behaviors data

The distribution of news history among users



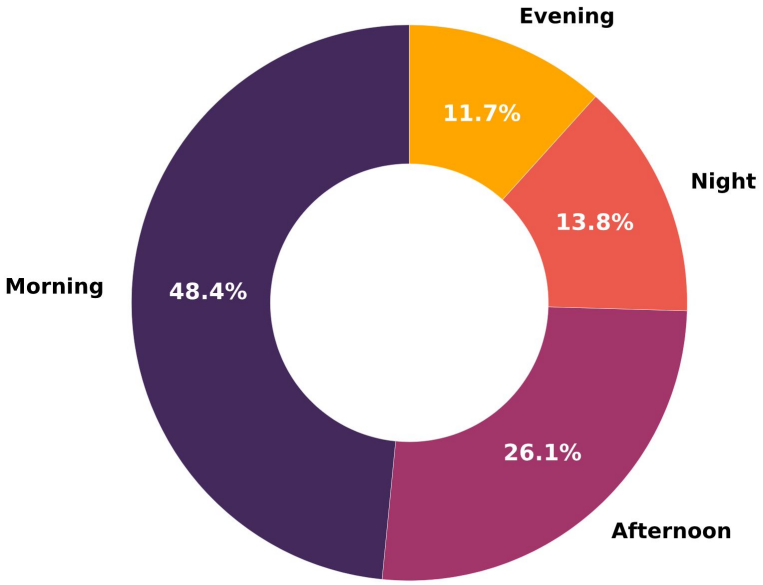
Outliers

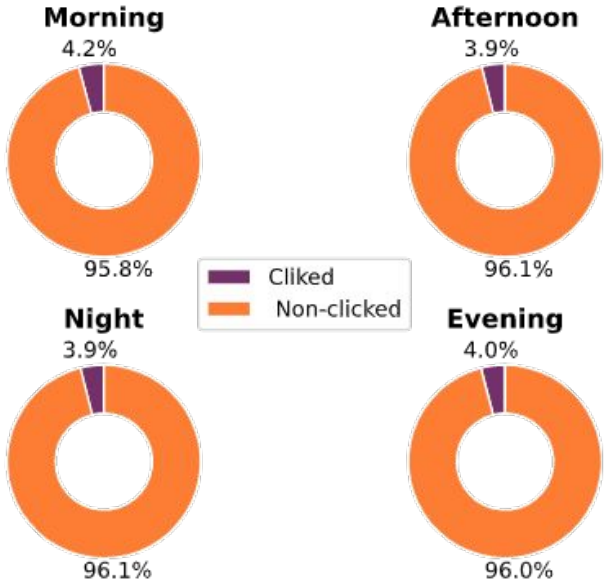
The data was taken during a 6-week period.

Some users start using the news platform earlier than others.

Time preferences for reading news

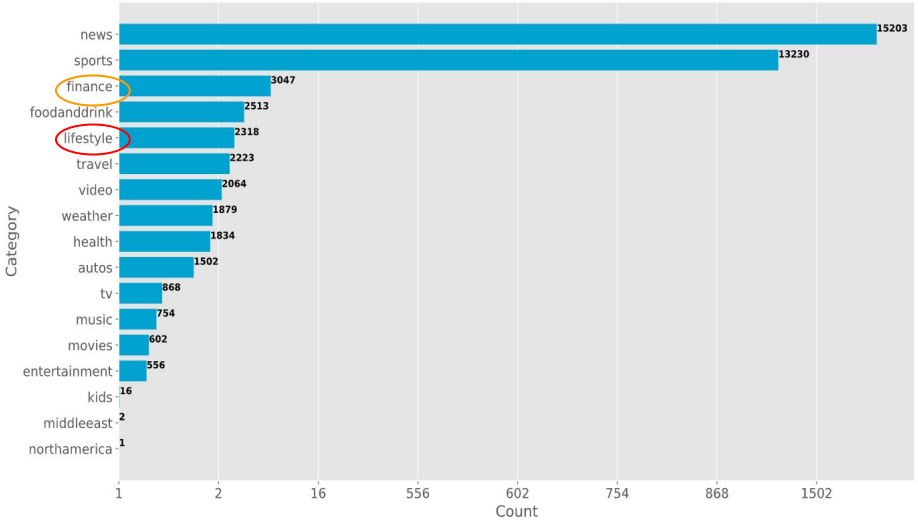
Most users prefer to read news in the morning **but** time does not affect the impression rate.



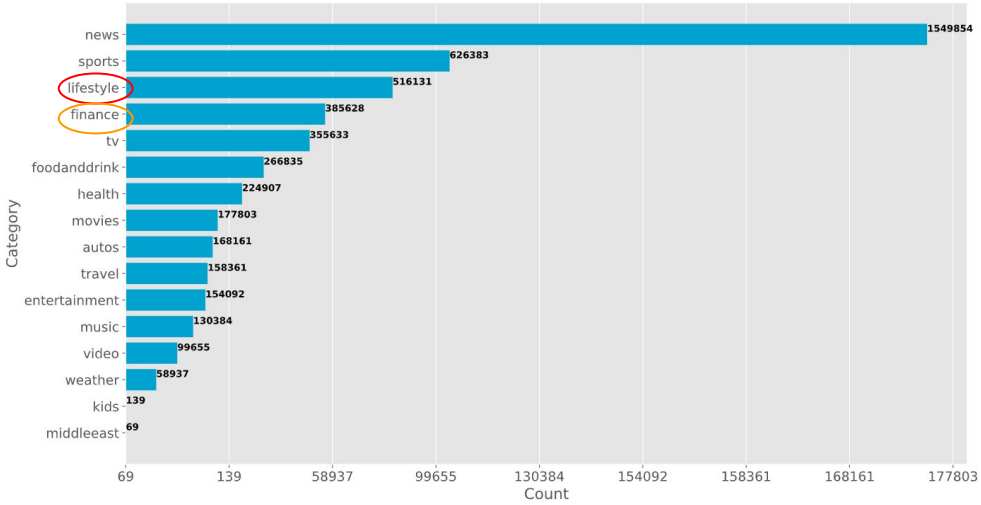


The popularity of news category

Based on all news collection



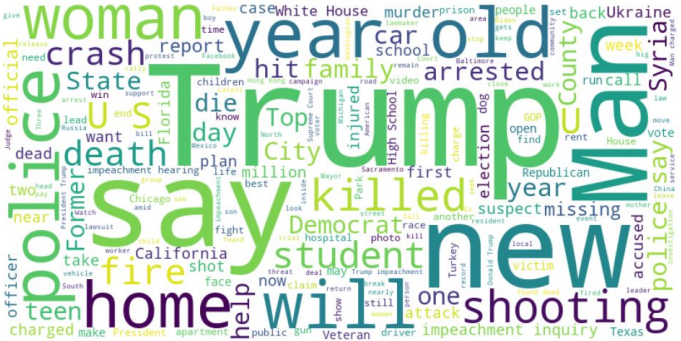
Based on users' history



Users like **lifestyle** category more than **finance**!

The popularity of news topics

Based on all news collection



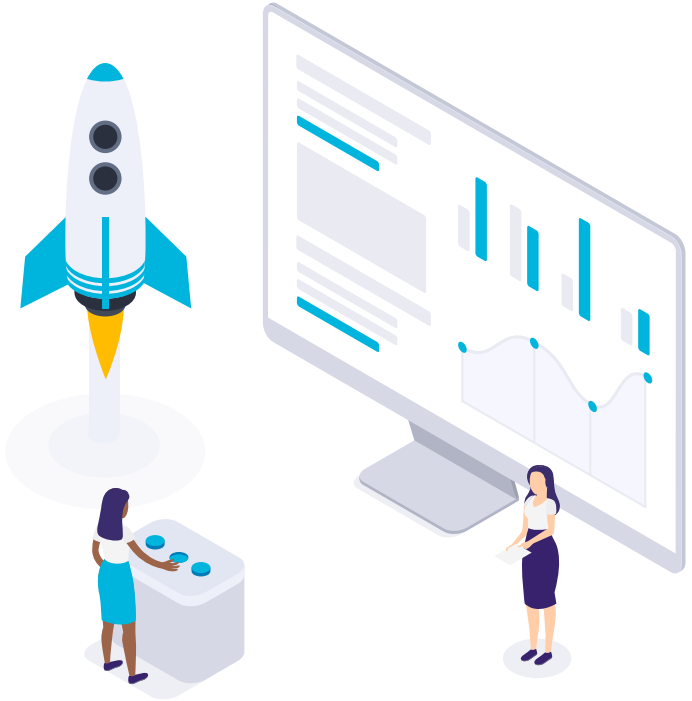
Based on users' history

Topic types: Celebrities, Organizations, Events, Timestamps (festivals) ,Geological Information

The most popular topic: Donald Trump*

*: This happened because there was **an impeachment inquiry against Donald Trump** during the dataset collection period (October 12-November 22, 2019).

News Recommendation System



Methodology

Task Definition

Input:

- News dataset
  Time, title, abstract, and news category
- User dataset
  Reading history, and preference of historical news

Task:

- Recommend news in dataset for users

Methodology

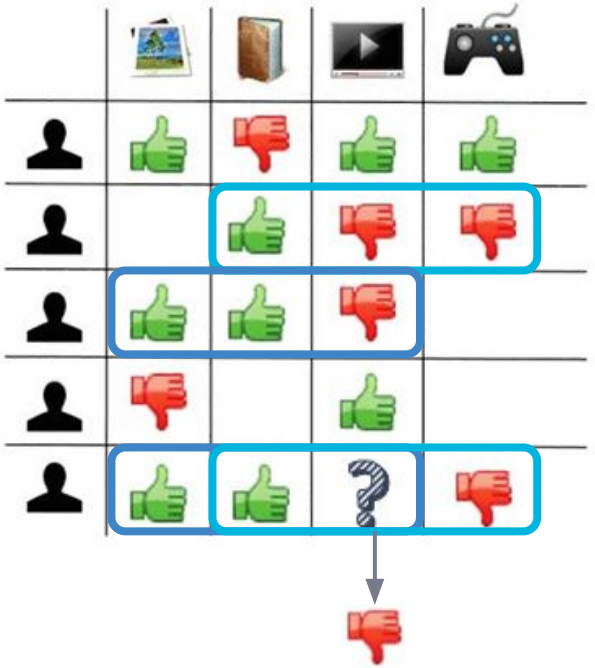
Baseline: Item-based Collaborative Filtering (Item-CF)

Main idea:

Recommendation based on similarity between items
Similarity is calculated using users' evaluation

Drawbacks:

- Poor performance on sparse datasets
- Missing semantic information



Methodology

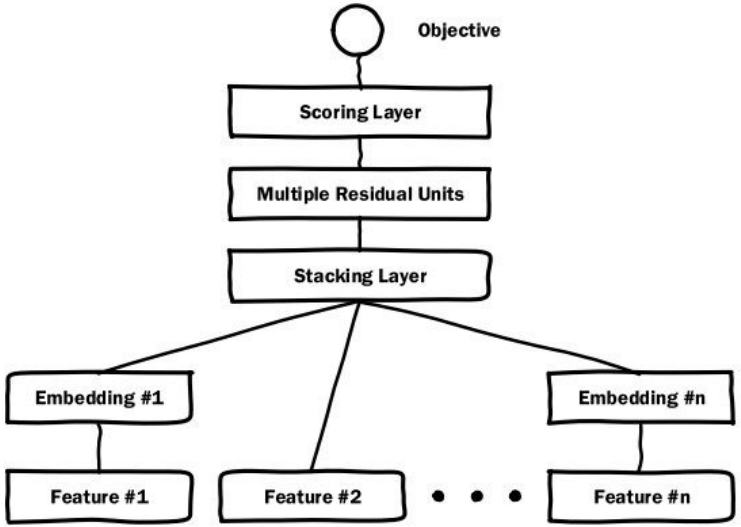
Baseline: Deep Crossing

Main idea:

Focus on embedding crossing preference of news

Drawbacks:

- Poor performance on sparse datasets
- Missing semantic information



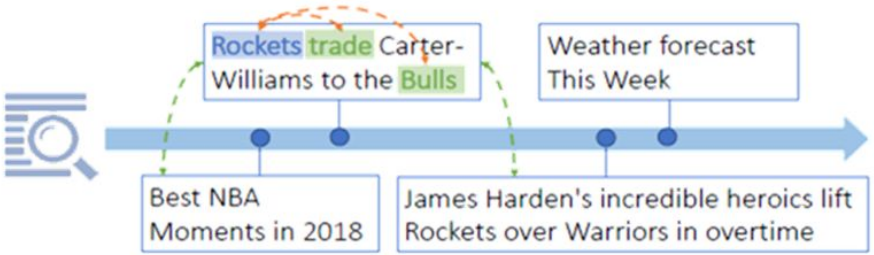
Methodology

NRMS: Neural News Recommendation with Multi-Head Self-Attention

Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., & Xie, X. (2019, November) [2]

Main idea:

- Model news presentation and user behavior simultaneously, then predict click
- Select important words with attention mechanism[3] to learn informative news representations



Methodology

NRMS: Detailed Model

Input:

- News title
- Abstract (Optional)
- Category (Optional)
- User history and preference

Label:

- Preferences of news (0/1 for dislike/like)

Output:

- The confidence to recommend a given news to a certain user

Hyperparameters	Value
Max Title length	10
Max Abstract length	50
Number of Multi-Head Attention	10
Dimension of pretrain GloVe Word Vectors	300
Negative Sampling K	4
Maximum number of historical News	50
Vocabulary Size	40000

Result

Recommendation result on MIND

Method	AUC	NDCG (First 5)	n-DCG (First 10)
Baseline	0.640		
NRMS: Title only	0.707	0.385	0.450
NRMS: Title + Abstract	0.714	0.410	0.465
NRMS: Title + Abstract + Category	0.724*	0.420	0.493

*: Achieved higher AUC than the best result(0.714) in the original paper

Result

NRMS: Sample of recommendation on unlabeled data

UserID = 90, number in brackets shows the times of appearances

Reading history:

Lifestyle(6), Music(3), News(3), Sports(2), TV(2), Movies(1)...

Noticeable topics:

- Royal family, celebrities & entertainment industry (12)
- Meghan Markle (5): Duchess of Sussex
- Wendy Williams (1): Broadcaster, hold TV show 'The view'

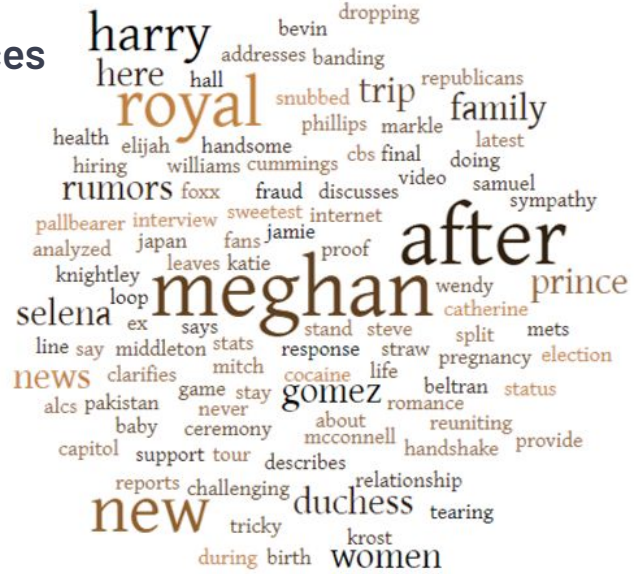


Fig: Word cloud of user's reading history

Result

NRMS: Sample of recommendation on unlabeled data

Green

Original keyword

Blue

Semantic association

Orange

Knowledge-based association

No.	Confidence	Category	Candidate News
1	0.869458	lifestyle	Meghan Markle and prince Harry won't spend Christmas with Queen Elizabeth at Sandringham this year
2	0.702128	tv	Kaley Cuoco says having separate lives has helped her marriage to Karl Cook
3	0.697480	entertainment	Camila Cabello meets with Kate Middleton & Prince William
4	0.691352	lifestyle	Meghan Markle chose a chic black ensemble for remembrance Sunday services
5	0.685914	tv	Whoopi Goldberg Addresses 'The View' Tension: 'If We Were Fighting, You'd Actually Know It'

Noticeable topics: Meghan Markle, Royals, celebrities, family, women

Conclusion

- NRMS achieved best recommendation result on MIND
- NRMS further passed verification on unlabeled news
- Future works may include long-term user behavior

References

[1] Wu, F., Qiao, Y., Chen, J.H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W. and Zhou, M., 2020, July. Mind: A large-scale dataset for news recommendation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 3597-3606).

[2] Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y. and Xie, X., 2019, November. Neural news recommendation with multi-head self-attention. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 6390-6395).

[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need. arXiv preprint arXiv:1706.03762.

THANKS!

Questions are welcomed