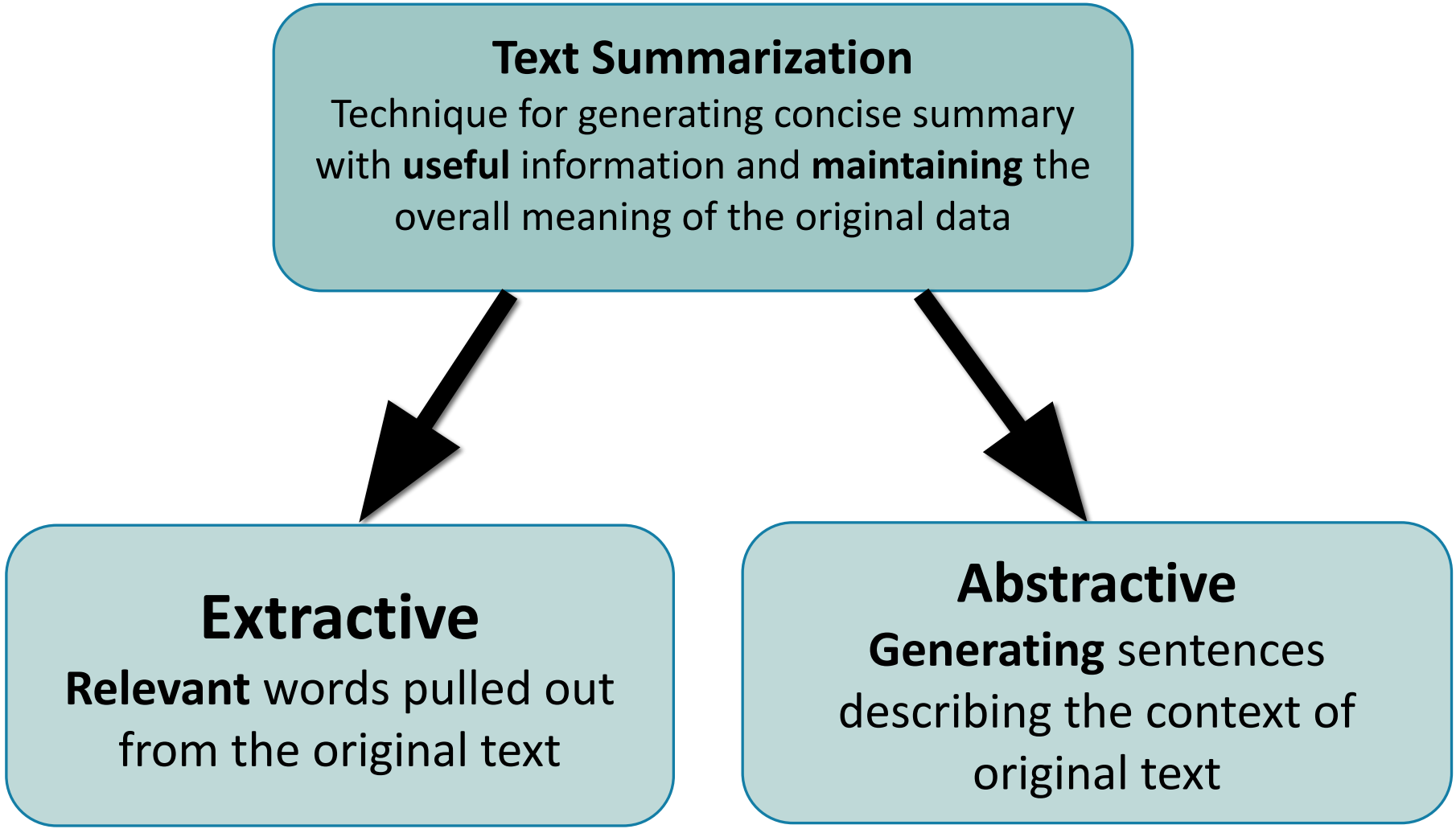
QUERY-BASED SUMMARIZATION

Manaswini Nagaraj, Juhi Paliwal, Viramya Shah

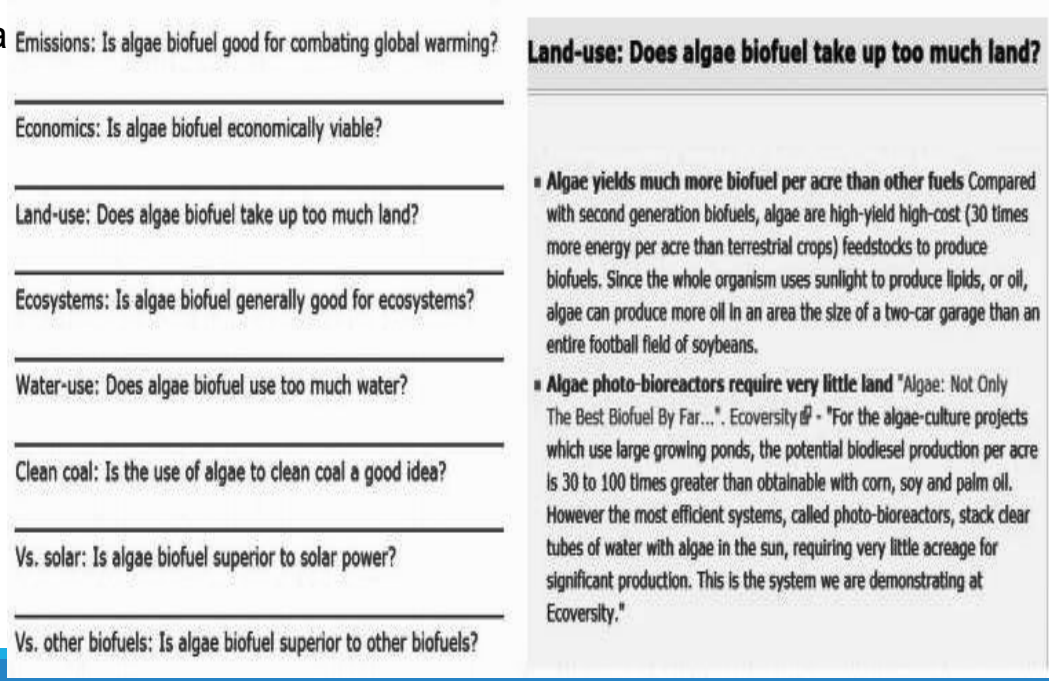

Text Summarization
Technique for generating concise summary with **useful** information and **maintaining** the overall meaning of the original data
Extractive
Relevant words pulled out from the original text
Abstractive
Generating sentences describing the context of original text

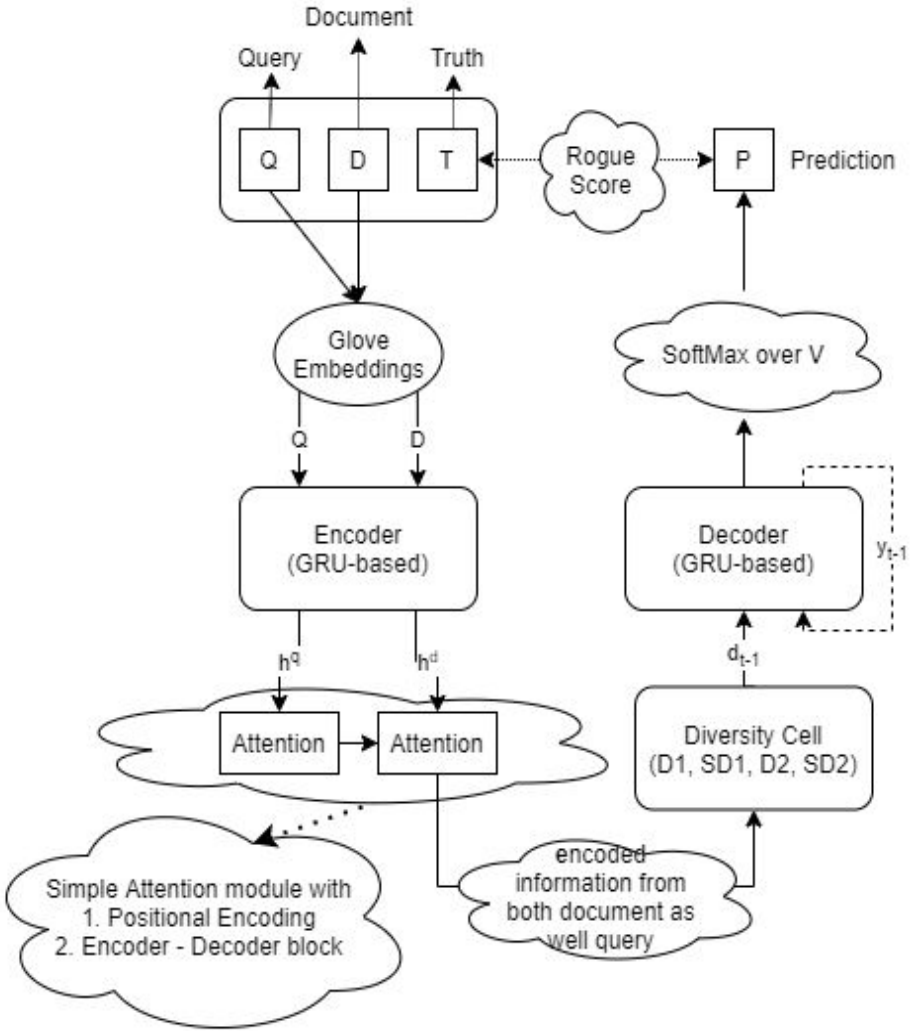
DATASET

A collection of corpus on various debate topics collected from Debatepedia

- **663 debates**, with documents which all have a least one query
- Debate = {**query,document,summary**}
- Avg queries/debate = 5
- Avg documents/query = 4
- The **average word** count per:

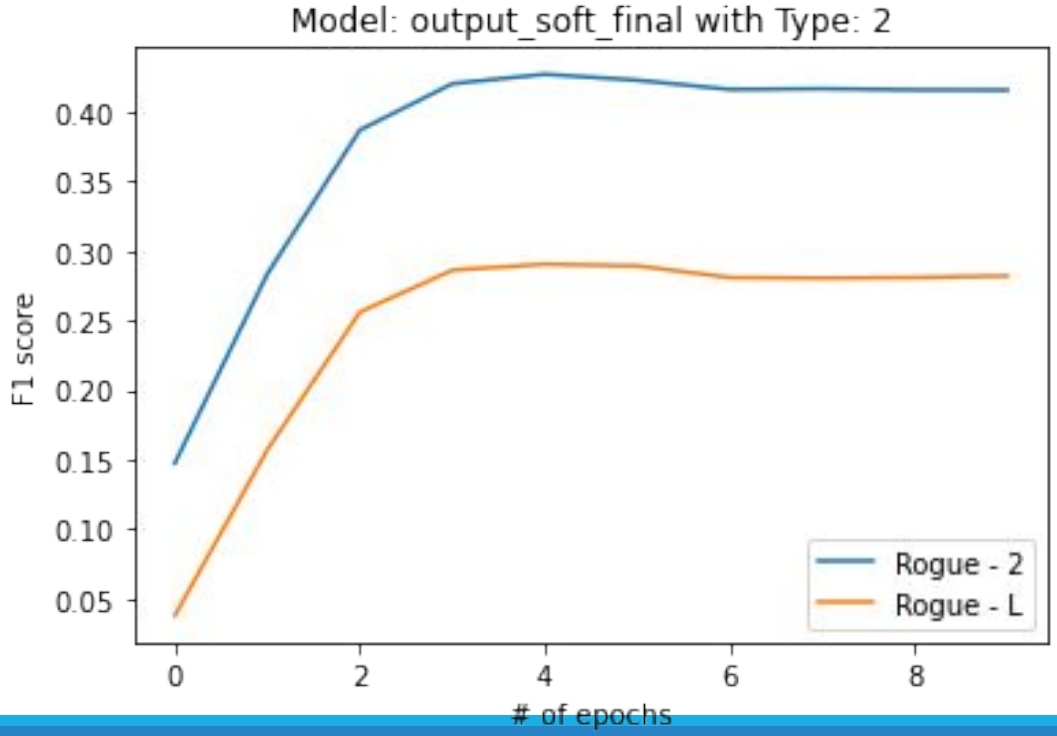
Document	Summary	Query
66	11	10





MODEL TECHNICALITIES

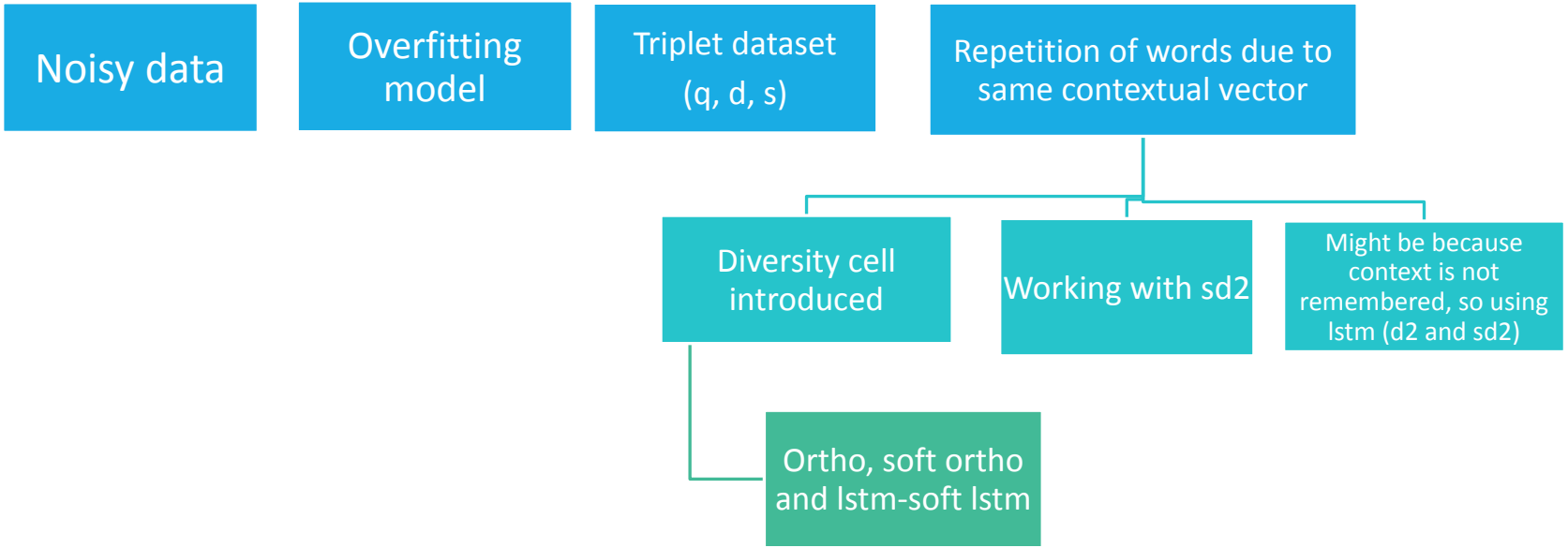
RESULT


Model: output_soft_final with Type: 2
F1 score
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
2
4
6
8
# of epochs
Rogue - 2
Rogue - L

PREDICTED SUMMARY

clean_true	clean_pred
alternatives to hydrogen exist to lower foreign oil dependencies	**alternative to hydrogen exist to lower overseas oil dependencies**
of powers in reactors is not economical	**nuclear electricity is expensive**
communism run by bureaucracy not the people	communism run by central bureaucratic not the people
the repercussions of such a bans want not be severe	the consequences of such a bans need not be severe
parents should talking to their children about healthier eating habits	parents should talk to their child about healthier eating habits
wikipedia s popular makes the spread of common	wikipedia s popularity makes spreading common
concerns over women s rights can not be blamed on	rights can not be blamed on
progressive taxation is socialist wealth redistribution and state	progressive socialism enable an making the tax
people need not see bull die in order to know	hunt is hunt for self over animals to their
wheat methanol energy produces a profit energy gain	electric cars can back-up potential out ethanol
parenting can making kid mentally	children s often make kid queen physically

ROADBLOCKS AND SOLUTION



Conclusion:

- A Query-based Abstractive Summarization model
- The model is a diversification mechanism based on successive orthogonalization.
- Provides diverse context vectors at successive time steps
- Pays attention to words repeatedly if need be later in the summary.
- Adding an attention mechanism on the query string gives best summaries.

Future Scope:

- A Context-based Abstractive Summarization model
- Context = Single word/ phrase/sentence/query
- Opinion generation

Thank you!

MASK UP & STAY SAFE!