NEWS SUMMARIZATION

Sarath Manoj,
Akshul Mittal,
Amit Gupta,
Devansh Shrestha

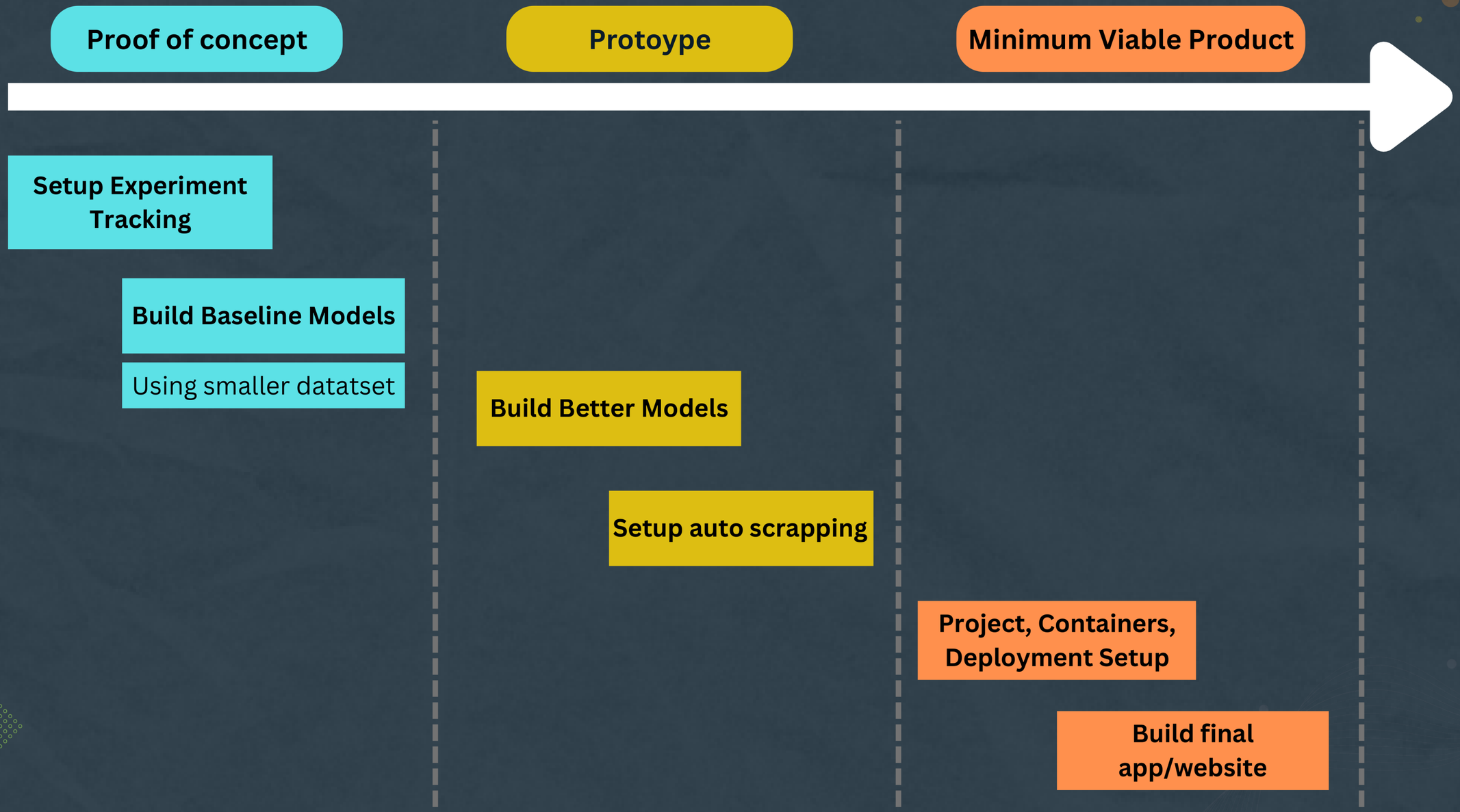
PROBLEM DEFINITION

Joseph wants to be updated on the current affairs to be an informed individual . But the large influx of information , shortened attention span and lack of time poses huge obstacles in this pursuit. Further, he is unable to consume all the wisdom that past literature works offer. Despite these challenges , he will continue to build his reading habits if they are more rewarding to match the more challenges he has to overcome.

Proposed Solution

We can lighten the burden of people like Joseph by providing meaningful summaries of news and other articles .
They can read from the selection of news summaries from scrapped news articles or input the article that they want to be summarized .
We do this by training an algorithm to identify the key insights and remove unnecessary description.

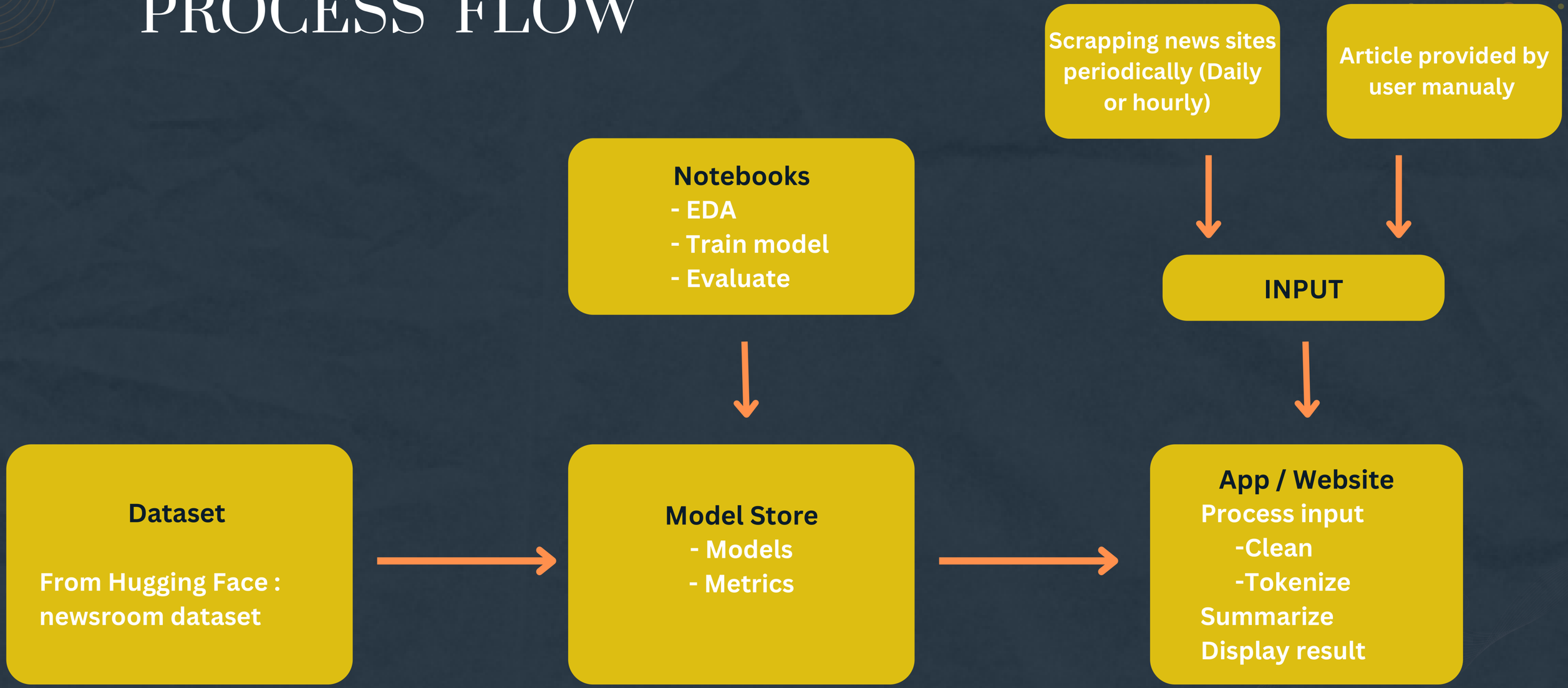
PROJECT WORKFLOW



TIME LINE

Proof of concept	Protoype	Minimum Viable Product
1 week	2 weeks	2 - 3 weeks

PROCESS FLOW



DATASET

CORNELL NEWSROOM is a large dataset for training and evaluating summarization systems. It contains 1.3 million articles and summaries written by authors and editors in the newsrooms of 38 major publications. The summaries are obtained from search and social metadata between 1998 and 2017

- Text
- Summary
- Compression
- Date
- Density
- URL
- Title

Dataset link: https://lil.nlp.cornell.edu/newsroom/index.html
https://huggingface.co/datasets/newsroom

MODELS

- Seq2seq
- Seq2seq + attention
- BERT
- Word2vec Embeddings
- Transformers