

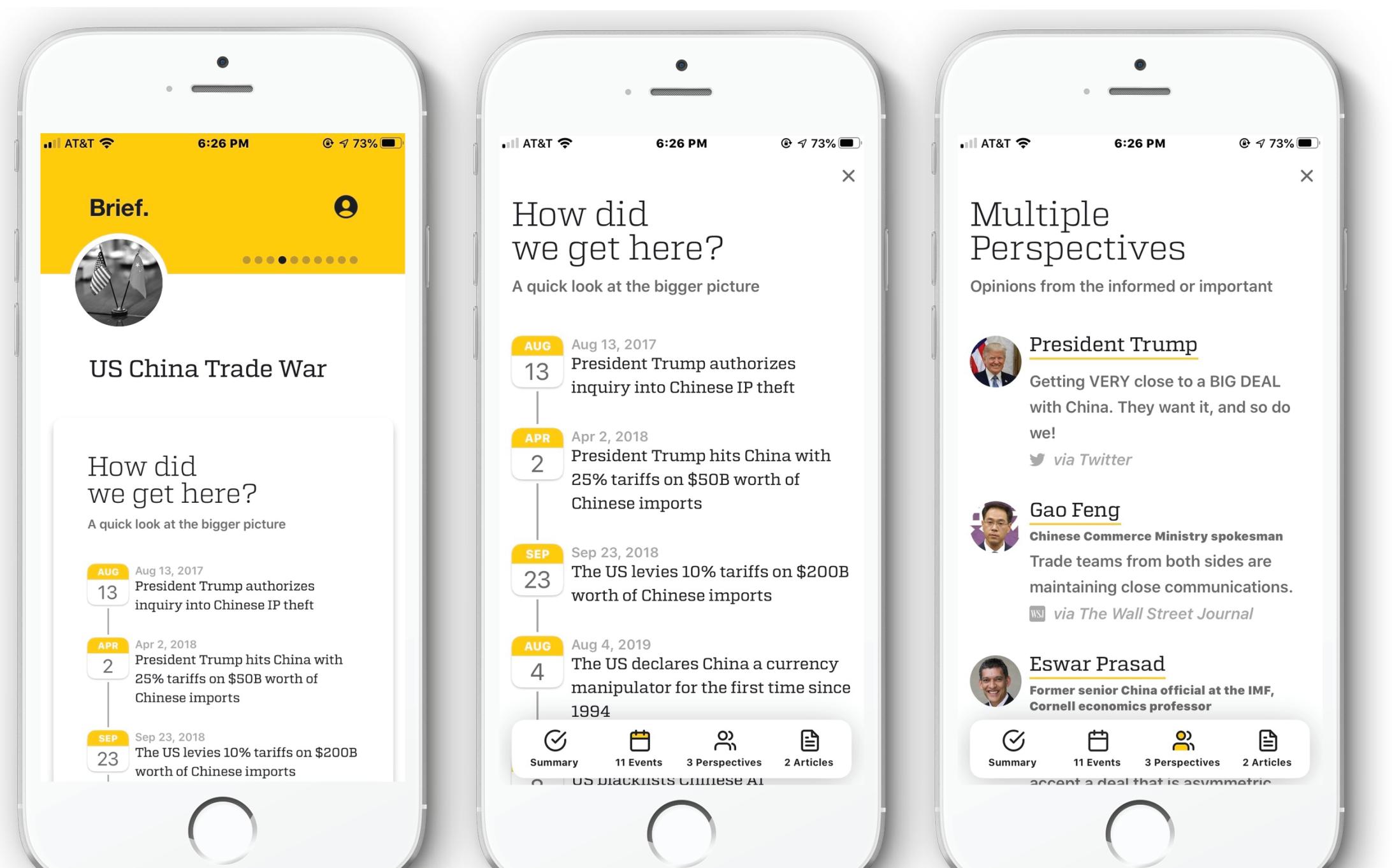
Purpose:

I explored the **ethical** considerations and **implementation** of machine learning as a tool for a news aggregator.

Context and Background:

The inspiration for this project is a new political news app that an Olin alumnus is building. I researched ways that machine learning could be applied to **assist in the editor's daily curation process**.

The App:



The app is a news aggregator that provides a summary of 5-8 news articles each day.

Stakeholders:

Behind the scenes, the team has an editor who compiles the topics for the day. It is a tedious process to look through 38 different news sources manually, so I explored the possibility of using LDA for Topic Analysis to assist the editor in this daily task. This model has the potential to **sort the 1000 daily articles into 20 unique topics** that the editor could choose to write about.

LDA Topic Analysis for News Aggregators

Raquel Dunoff

The Model:

Linear Discriminant Analysis (LDA) is an algorithm that uses word frequency to cluster news articles into common topics.

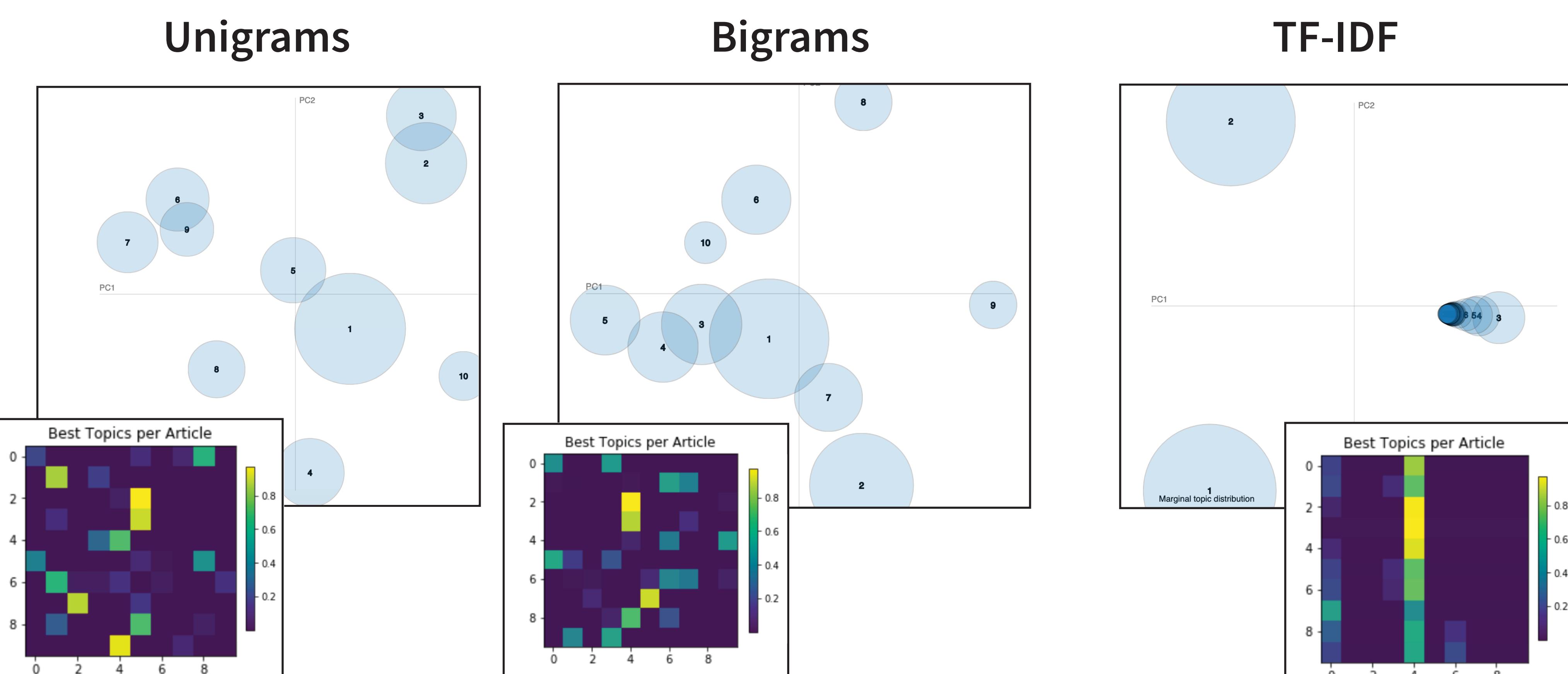
1000 Articles / Day → 20 Unique topics

Dataset:



All The News dataset from Kaggle. Over 100,000 articles from 15 different sources!

Results:



Unigrams and bigrams are both effective vectorizers for Linear Discriminant Analysis

Ethical Implications:

Algorithms limits the editor to only reading articles within the dataset of the same 38 news sources. This **silences the perspectives of local news**. It is important that the editors understand that the algorithm is only a **supportive tool** and that their judgment is essential to ensuring quality and truthful reporting.

Identifying Sources:

When aggregating the news and providing summaries, it is also important to **prominently identify the source** of the articles. This gives credit to the original journalist and allows the reader to discover who the author and publisher are, in order to **navigate biases**.

Social Impact Statement:

According to the FAT/ML organization, a Social Impact Statement should be created and reassessed during the design stage, pre-launch, and post-launch.

The sections of a Social Impact Statement are:

Responsibility **Fairness**

Explainability

Accuracy **Auditing**

Learn More:

Ask me about the contents of each of these sections for the app's Social Impact Statement.