THE UNIVERSITY OF CHICAGO in collaboration with abbvie presents….

Automating Supply Chain Risk Management

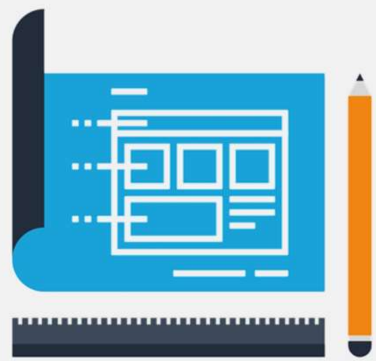
Created By:
Ji Hun James Lee

Agenda

1 Problem Statement

2 Dashboard Demo

3 Architecture



4 Crawler

5 Tagging Rules

6 Shared Services and Risk Models

7 Conclusions

8 Appendix

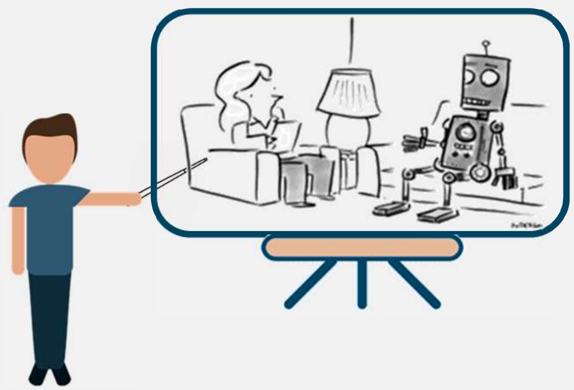
Problem Statement

Current State

Multinational pharmaceutical companies such as AbbVie have **large and complex supply chains**, spanning many countries and countless vendors.

Monitoring supply chain risk can be a **tedious, inefficient and time-consuming process for the staff**, although imperative for company's smooth functioning

AI Intervention



In response, we've developed an **AI-based solution automating the risk management process**

The risk rating system ingests data from news sources, identifies risky articles and summarizes risks in **three key categories: regulatory, socioeconomic, and geopolitical risk.**

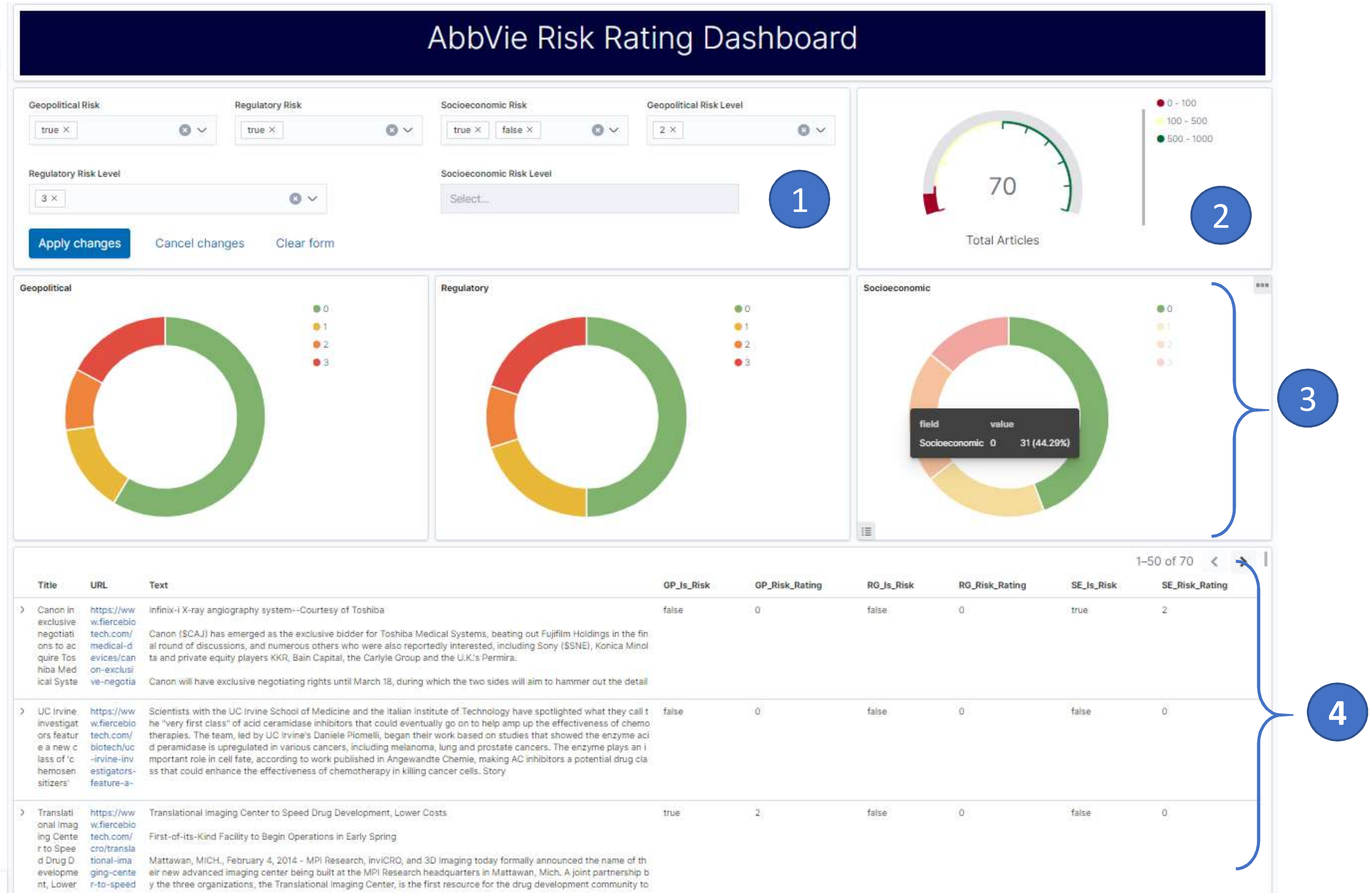
Future State

As a result, the risk management system **performs the bulk of the manual work** with limited manual intervention required from the staff.

The solutions also extracts key information from large texts such as **location of impact, supplier name** etc. for faster and informed decision making

Snapshot of the dashboard

Note: Data in the dashboard is simulated to maintain confidentiality.

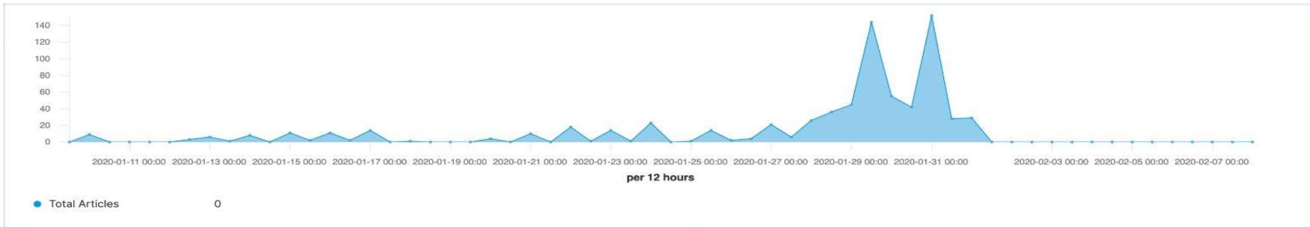


Components:

1. **Risk paradigm filter –** Choose geopolitical, regulatory, or socioeconomic focus
2. **Article count analysis –** The number of articles available for analysis within timeframe
3. **Risk level split –** Proportion of articles across each level for each paradigm (interactive)
4. **Table of articles –** Actual articles analyzed (interactive)

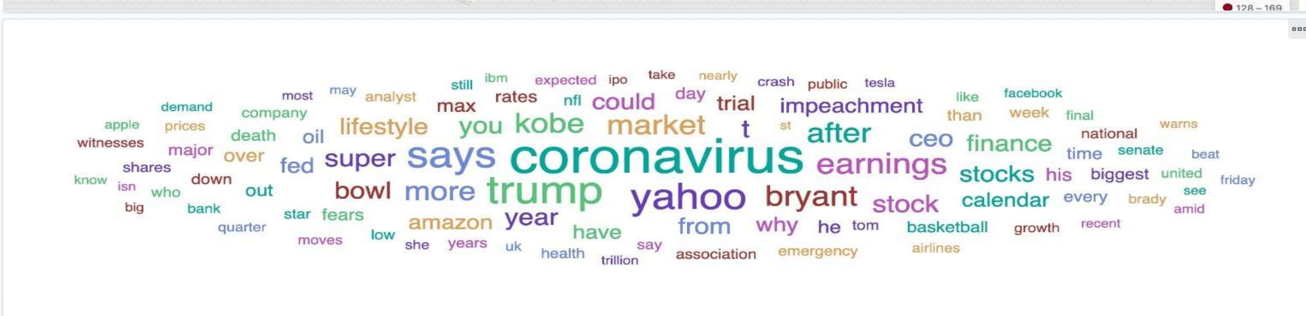
Snapshot of the dashboard (contd.)

Note: Data in the dashboard is simulated to maintain confidentiality.



5

6

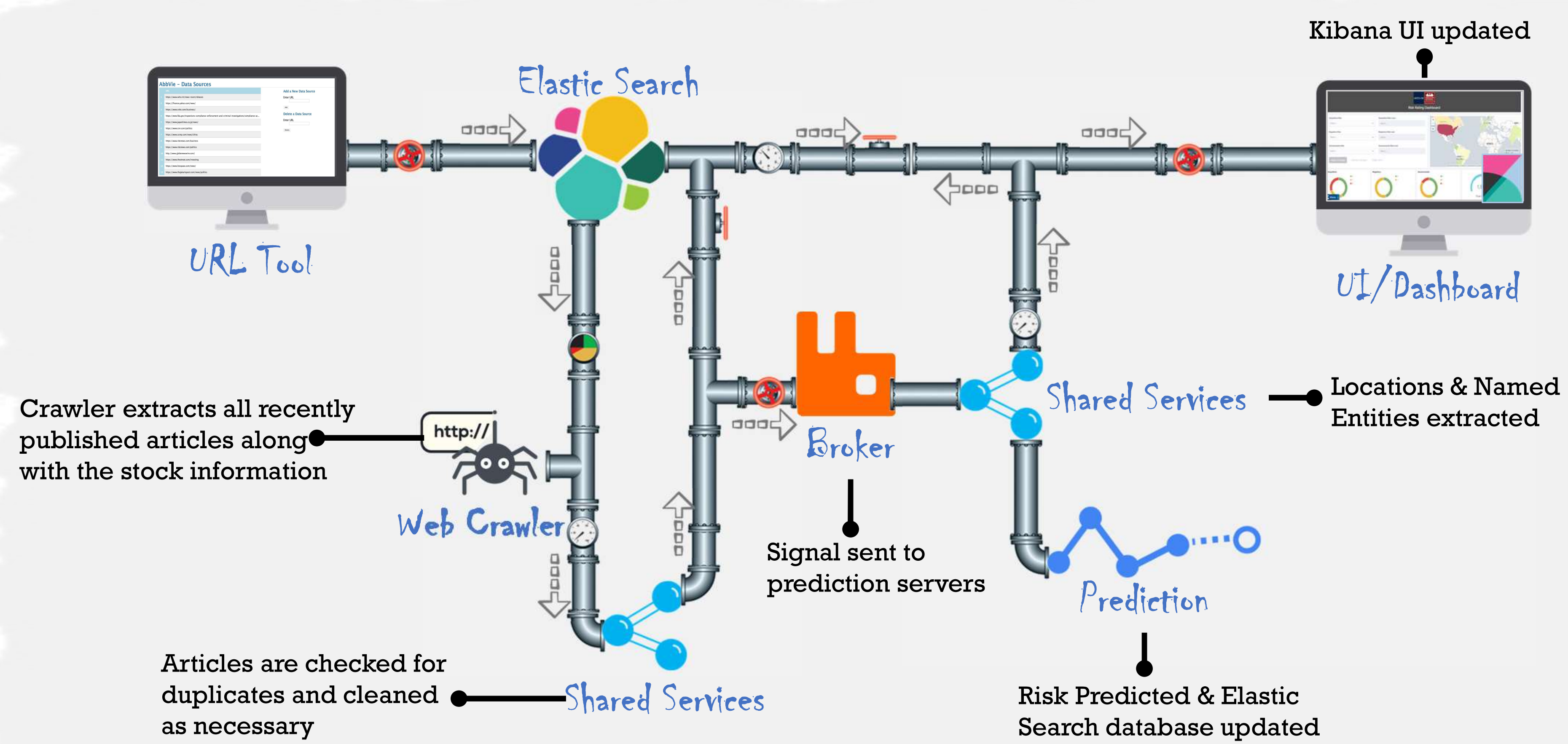


7

Components:

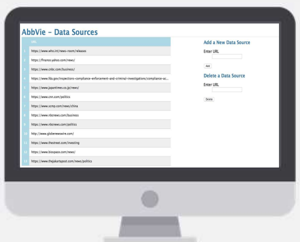
5. **Time series –** Plot on frequency of articles based on topic. For example, *coronavirus* can be used as a filter to visualize the number of articles mentioning it over time (interactive)
6. **Geographical distribution –** Geospatial layout of article distribution for topic (interactive)
7. **Wordcloud filter–** Wordcloud to highlight the most discussed topics with applied filters (interactive)

Architecture



Web Crawler – Key Components

Our web crawler extracts articles from **35+ websites daily,** along with stock market data from user-defined list of sources



Front End URL Tool enables **user to input custom list of news sources** for the crawler

Version 1.0 crawls website search engines using **keywords**

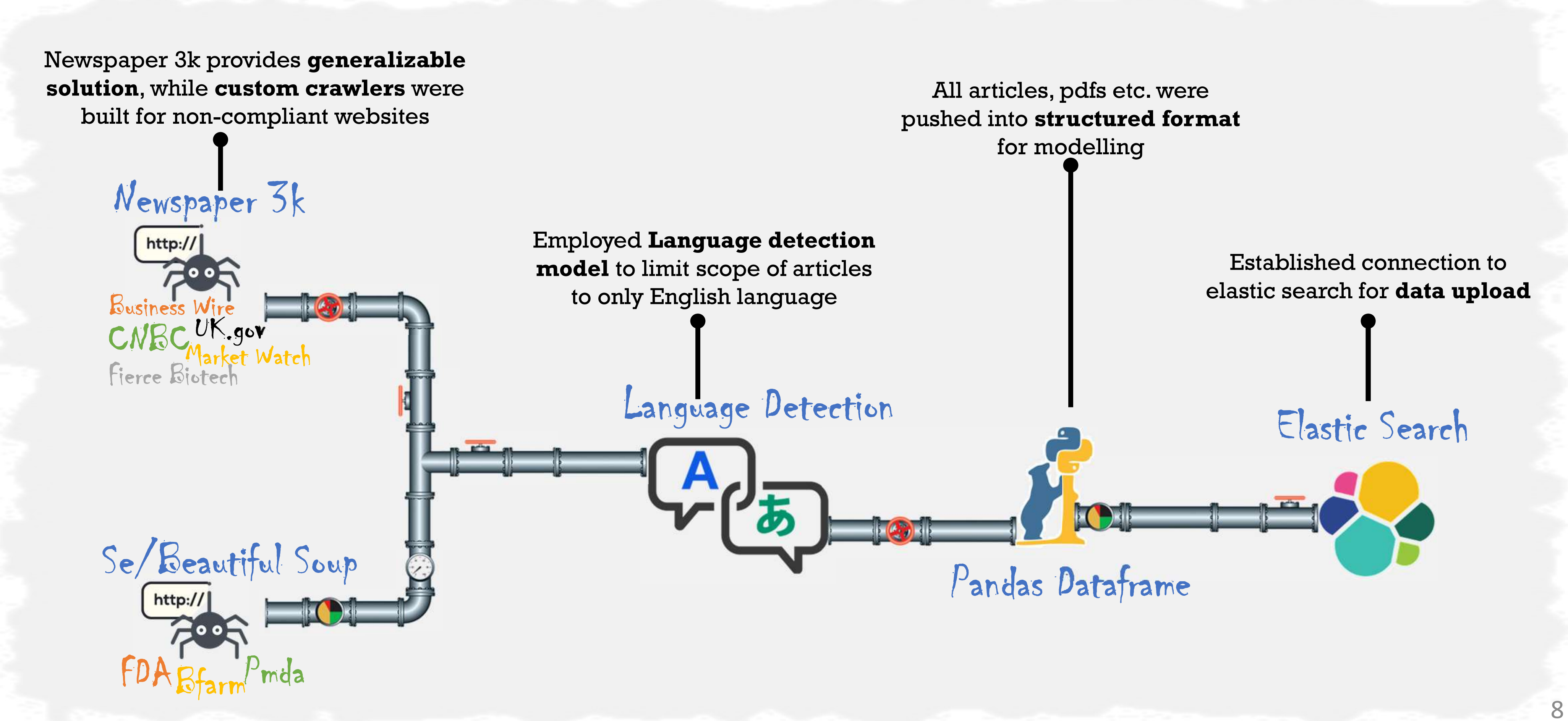
Title Text URL Publish Date Scrape Date

Version 2.0 of the crawler pulls out all articles within **website directories**

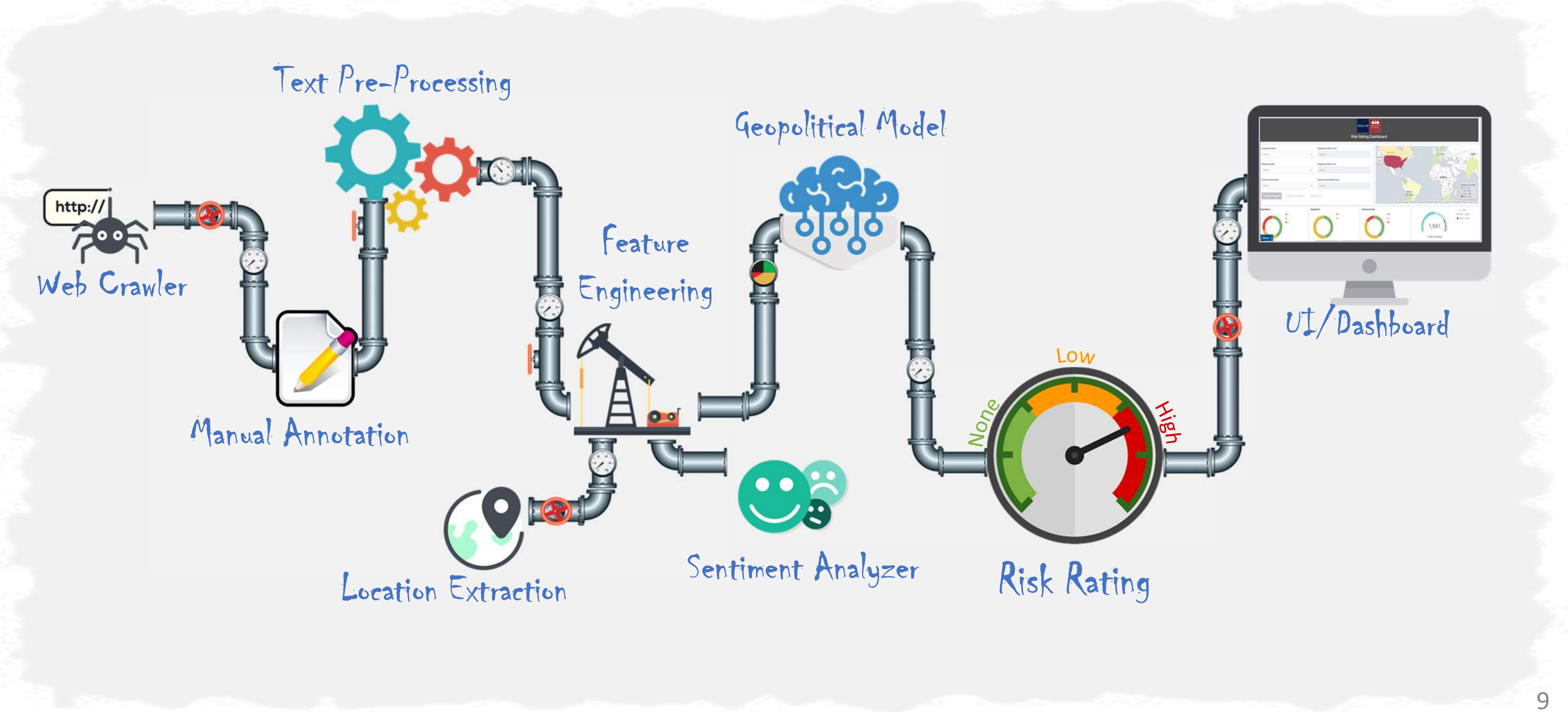
Extracts Stock Market Data for risk modelling

AbbVie Change
XBI Change
S&P Change

Web Crawler – Process Flow



Location Extraction and Geopolitical Risk Pipeline



Location Extraction

Data Collection

1. 500 articles

2. On average about 3 ***unique*** locations per article – 1,500 rows

3. Manually labeled as either ***'Correct'*** location or ***'Wrong'*** location. A location is correct when it is where risk event transpires.

4. Response variable is binary, hence ***Classification Problem***.. What are the features?

A U.S. GPE airstrike in Somalia GPE killed two civilians in 2018, U.S. Africa Command said in a press release on Friday, adding that it believed it to be an "isolated occurrence".

The April 1 strike near El Burr GPE , Somalia GPE , killed two civilians and four militants, not five militants as the U.S. GPE military had originally reported, a commander-directed review triggered by pressure from Congress and Amnesty International found.

The command has carried out 28 airstrikes in Somalia GPE in 2019, compared with 47 in 2018 and 35 in 2017. Most publicised strikes are followed with a statement on the number of militants believed killed.

Python's NLP module – SpaCy – has entity recognition function based on statistical model that identifies cities, states, and countries like above.

Location Extraction

Feature Extraction

I manually created features that could be predictive of the true location mentioned in article.

New Variables:

1. Proportion – relative frequency of location in article
2. First – whether location is mentioned first in article
3. Last – whether location is mentioned last in article
4. Mode – whether location is the most frequently mentioned in article
5. Title – whether location is mentioned in the article's title

View of Training Set

country.	proportion	first	last	mode	title
correct	0.5000000	1	0	1	1
wrong	0.0500000	0	1	0	0
wrong	0.0500000	0	0	0	0
wrong	0.0500000	0	0	0	0
wrong	0.3500000	0	0	0	0
correct	0.8666667	0	1	1	0
wrong	0.1333333	1	0	0	0
correct	1.0000000	1	1	1	0
correct	0.6000000	1	0	1	0
wrong	0.1000000	0	0	0	0

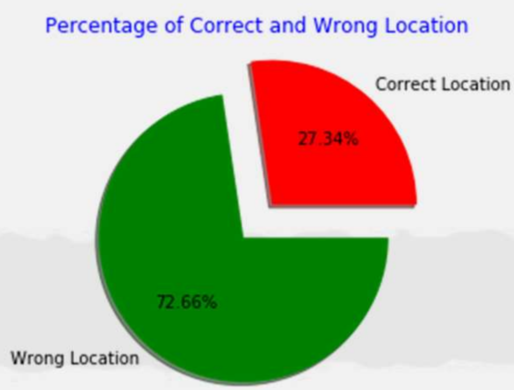
Every city and state was converted country level. For example, NYC is transformed to United States. On the other hand, all continent-level entities were removed.

Location Extraction

Exploratory Data Analysis

Response Imbalance

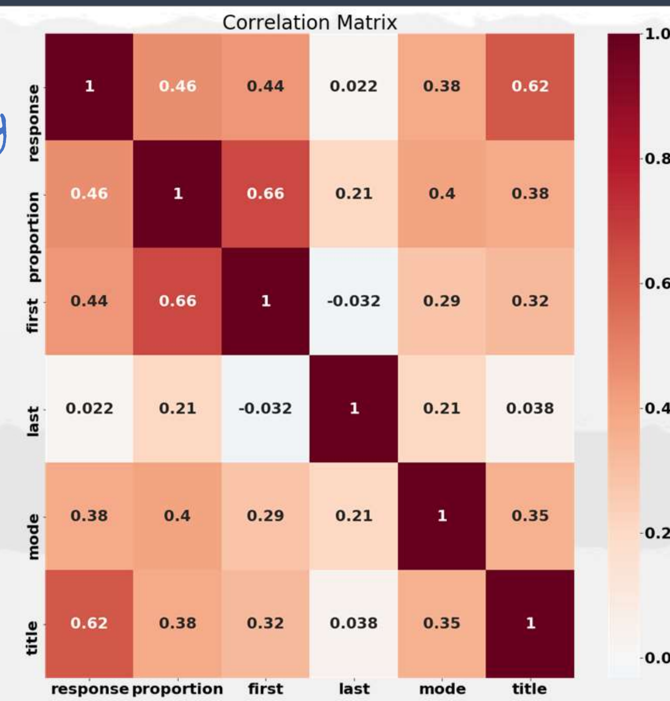
SMOTE to oversample Correct Location



Missing Value

No missing values, so no need for data deletion or imputation.

Multicollinearity



Outlier

Spacy made a few wrong classifications – for instance, Facebook and United Nations as locations.

Feature engineering function made bugs and created 100% proportion but non modal rows. They were deleted or modified.

Location Extraction

Feature Selection Methods

Set a default model

Compare feature subsets

Choose the final predictor set

Feature Selection Techniques	Proportion	Mode	Title	First	Last
Filtering (Chi-2 Test, One-Way Anova)	o	o	o	o	x
Recursive Feature Elimination	o	o	o	o	x
Simulated Annealing	o	o	o	o	o
Genetic Algorithm	o	o	o	o	o

O included X excluded

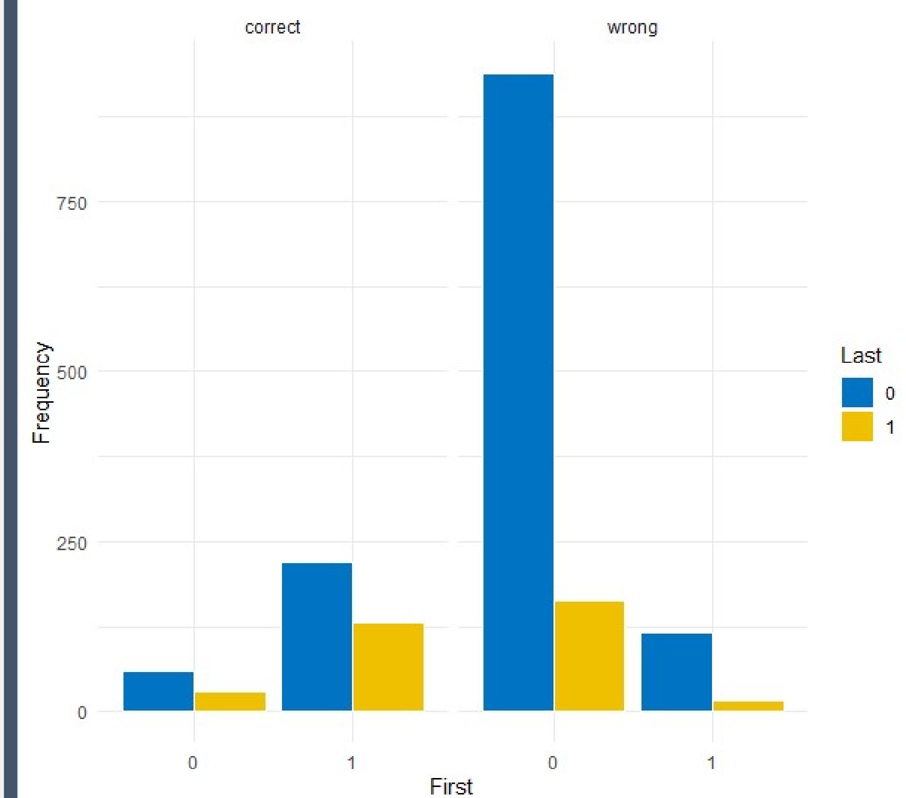
Location Extraction

Interaction Analysis

Logistic Regression Results

	Est.	S.E.	z val.	p
(Intercept)	-5.33	0.42	-12.80	0.00
proportion	0.67	0.17	3.82	0.00
first	1.97	0.48	4.08	0.00
last	-0.90	0.62	-1.45	0.15
mode	1.65	0.37	4.44	0.00
title	3.74	0.43	8.64	0.00
first:last	2.78	0.81	3.41	0.00
last:title	2.50	0.81	3.11	0.00

First vs Last vs Response Variable

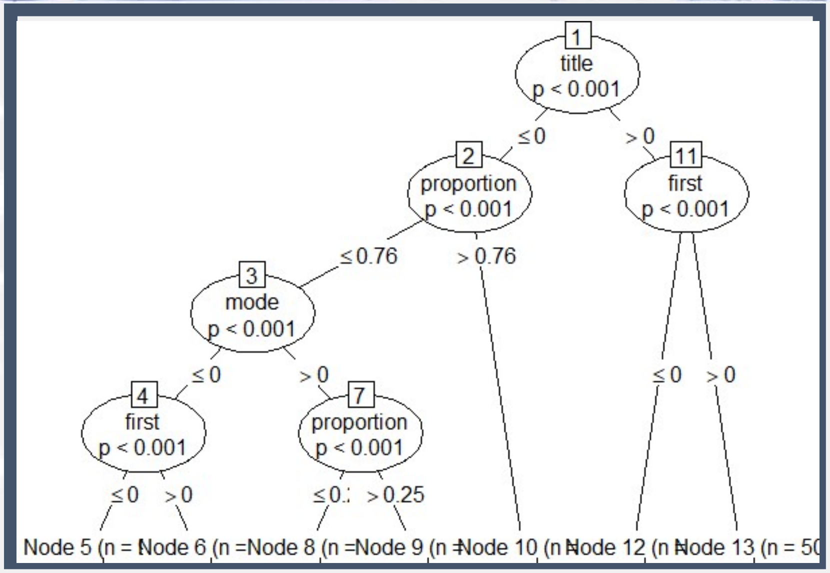


Location Extraction

Top Model with Grid Search

1. **Gradient Boosting Machine**
2. Conditional Random Forest

What is Conditional inference Tree?



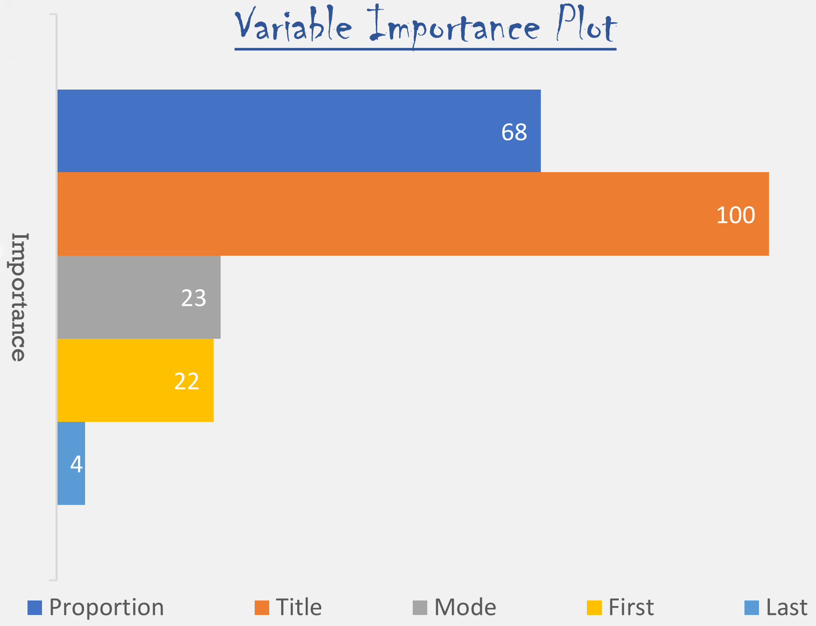
Evaluation Metrics

	Conditional Random Forest	Gradient Boosting Machine
Accuracy	0.981	0.977
Kappa	0.811	0.811
AUC	<u>0.878</u>	<u>0.913</u>
Sensitivity	0.994	0.986
Specificity	<u>0.762</u>	<u>0.840</u>
Pos Pred Value	0.986	0.989
Neg Pred Value	0.889	0.808
Precision	0.986	0.989
Recall	0.994	0.986
F1	0.990	0.988
Balanced Accuracy	<u>0.878</u>	<u>0.913</u>

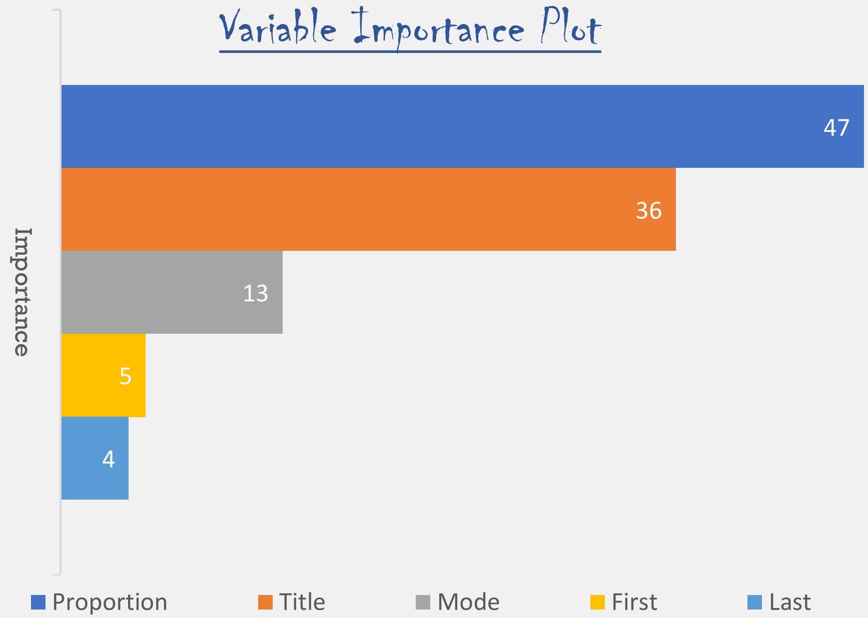
Location Extraction

Variable Importance

Gradient Boosting Machine Variable Importance Plot

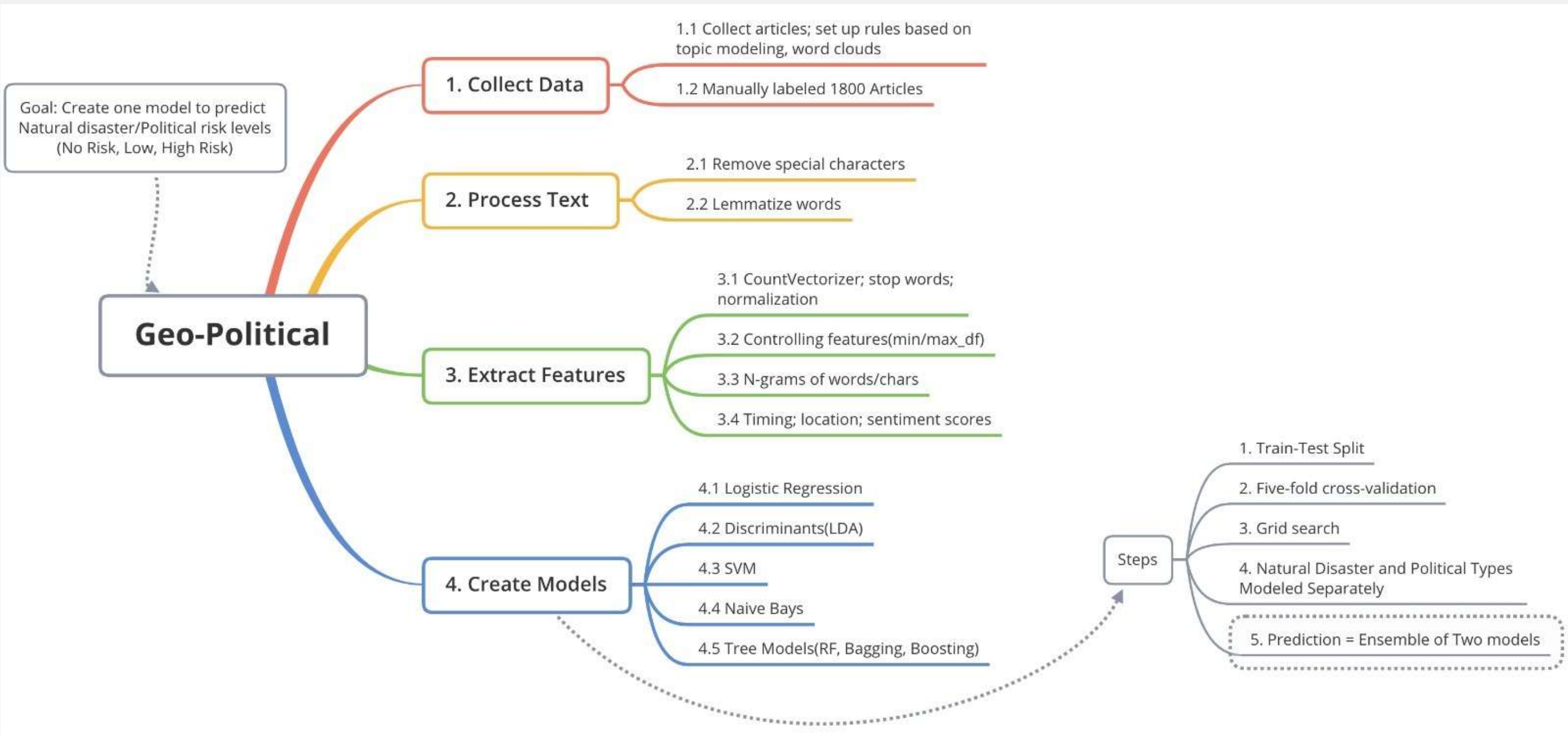


Conditional Random Forest Variable Importance Plot



Geopolitical Risk

Model Summary



Geopolitical Risk – Annotation Rules

Natural Disaster

No Risk Historical and fictitious events; personal disaster (E.g.: house fire); far from human civilization

Low Risk Small span of region; few casualties; insignificant damage to property

High Risk Large span of region; high number of casualties; severe property damage; extended periods of time

Political Issues

No Risk Active war but peace treaty/truce/armistice; historical and fictitious events

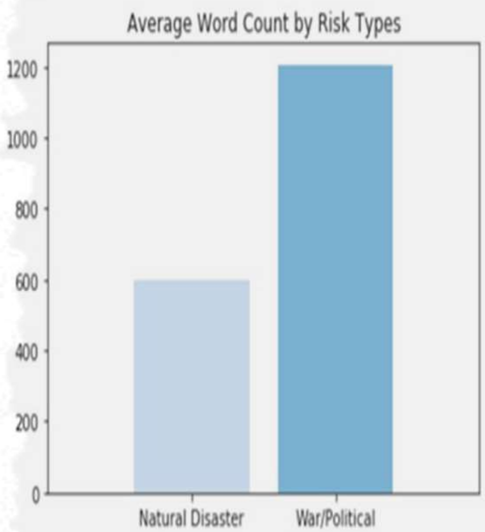
Low Risk Terrorism - explosion or violence (no death toll), internal conflicts with low death toll

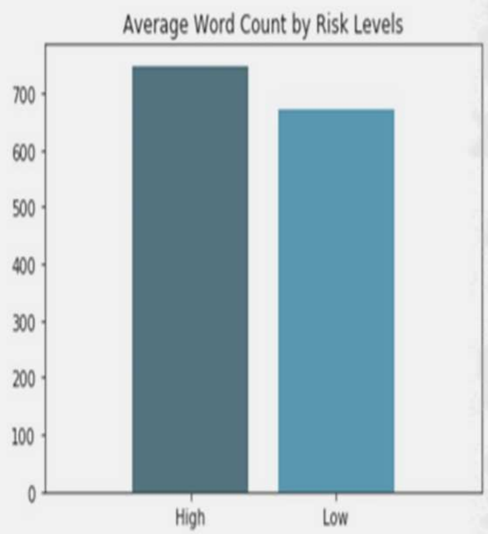
High Risk War – airstrike and deaths; military conflict between nations; nuclear proliferation; terrorism with high death toll

Geopolitical Risk – Exploratory Data Analysis

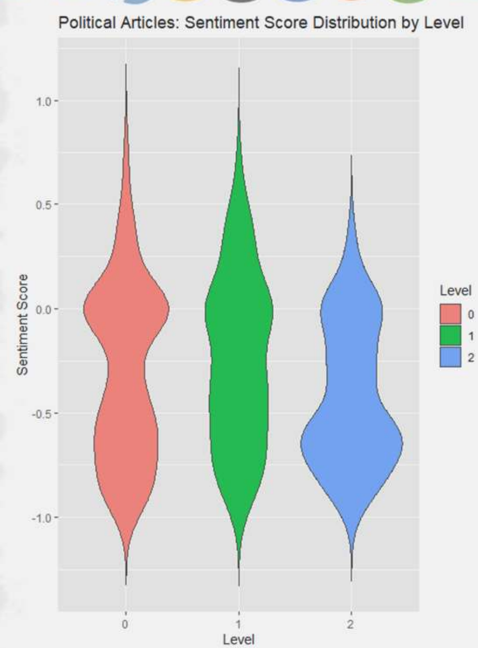
Interesting Findings on Word Count and Sentiment Analysis

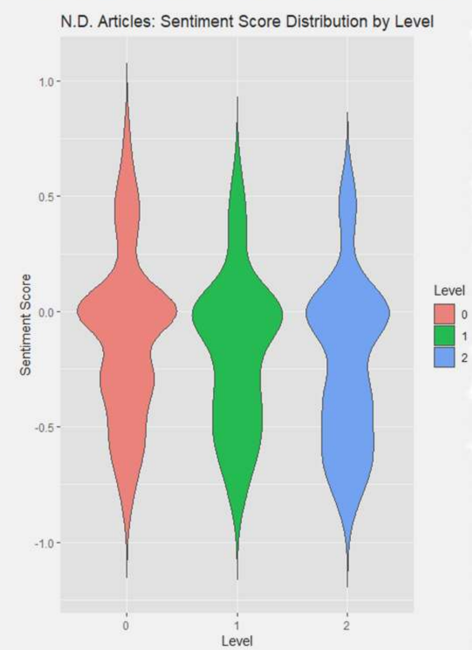
Word Count





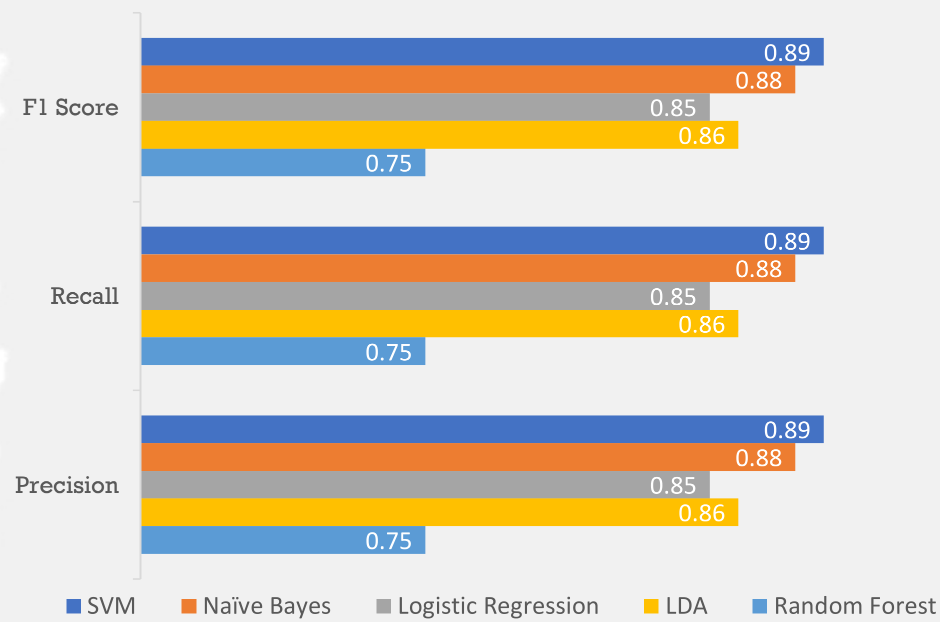
Sentiment Analysis





Geopolitical Risk – Model Results

Models Performance Comparison



SVM performs the best among other models, with 0.89 precision and recall for all risk types(Political/Natural) combined

Final Model Results

SVM

Naïve Bayes

Employing individual models for Natural Disasters and Political risk types improved model performance. Final models(both) yielded the following results:

Precision	Recall	F1 Score
0.90	0.89	0.90

Regulatory Risk

What constitutes a Regulatory Risk?

Articles 1 and 2 are both flagged as risk by our **supervised ML model** based on manual annotation of ~1200 articles.

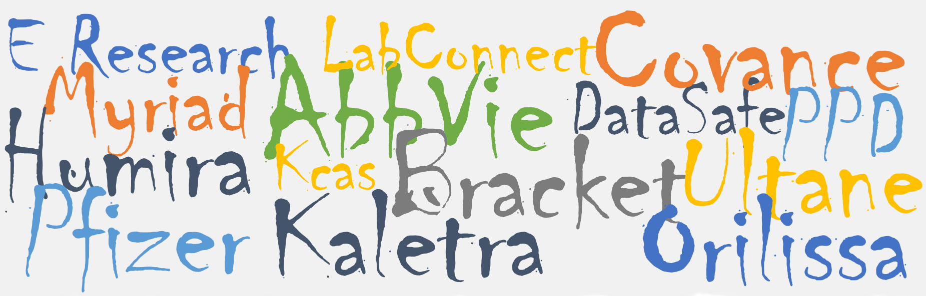
However, the **Keyword Detection** model identifies vendor name and flags only article 2 as 'Risk' and the other as 'Non-risk' based on relevancy to AbbVie.

What sources were considered?

Our source list is focused towards Regulatory Agency websites that govern and audit the pharmaceutical industry around the globe.

What keywords were considered?

Our custom fuzzy match algorithm looks up the authorized list of AbbVie's vendors and drugs associated with AbbVie.



Regulatory Risk - Methodology

Socio-Economic Risk

Overview

AbbVie partners with vendors and corporations across the globe to accomplish its mission to treat and cure diseases and assist as many people as possible.

Distinguishing the socio-economic issues that may negatively impact AbbVie's vendor partnerships could limit the firm's risk exposure in its partnerships.

Stock Index

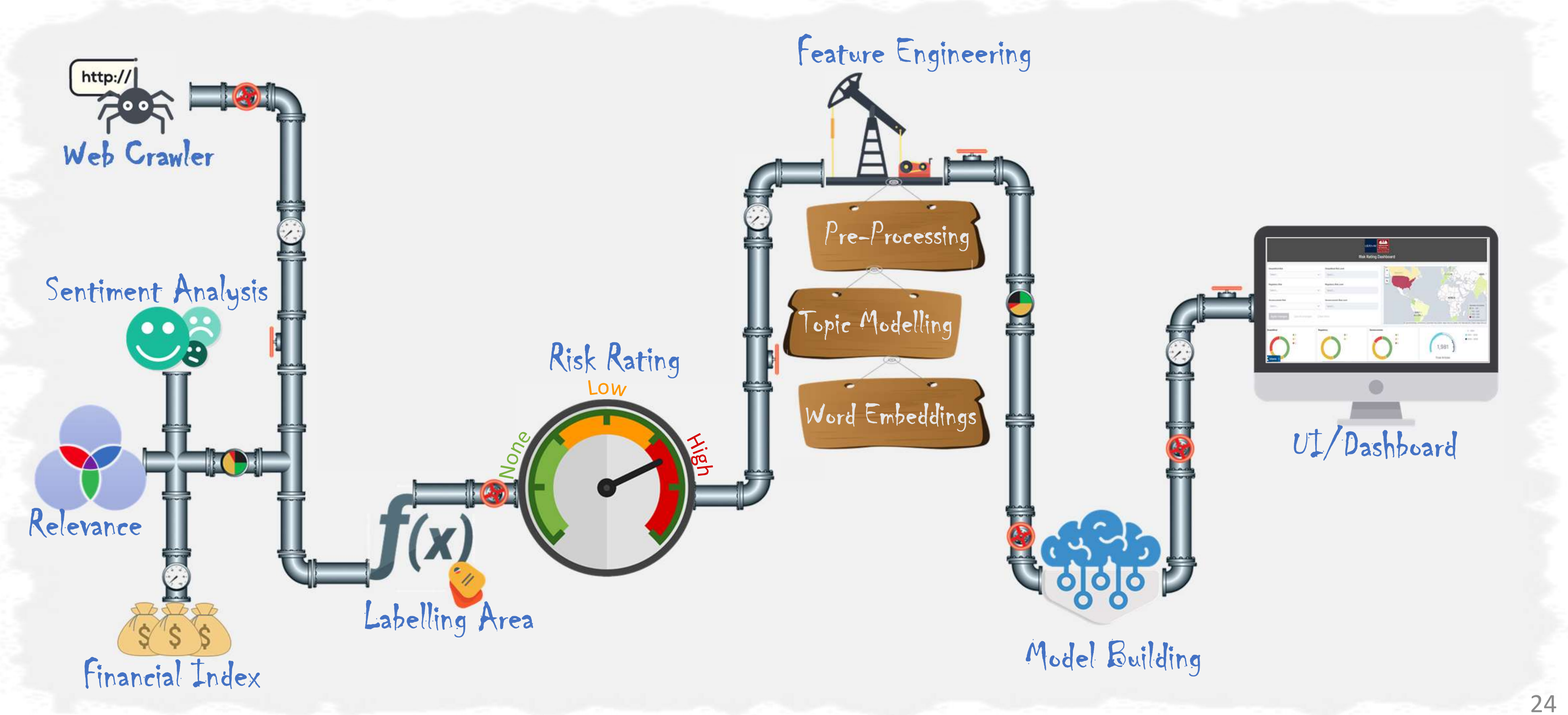
During our research we discovered that the **XBI** and **S&P500** spread was a good indicator to economic risk and combined with the sentiment of the news article would provide a pivotal foundation for our analysis

Three Pillars

We used three critical pillars in our analysis and final solution

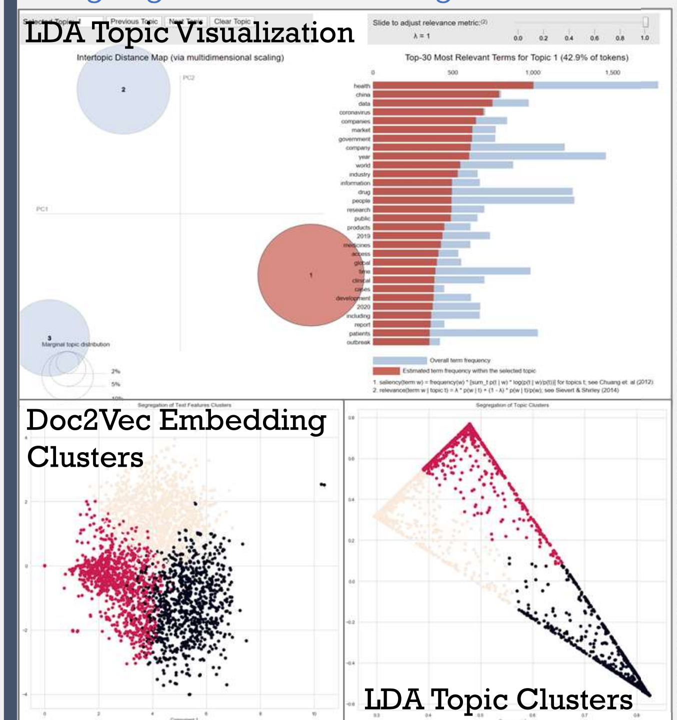
1. AbbVie team identified web sources for article data collection
2. AbbVie vendors and partners list
3. External financial data for the XBI, S&P500 and AbbVie index

Socio-Economic Risk - Pipeline

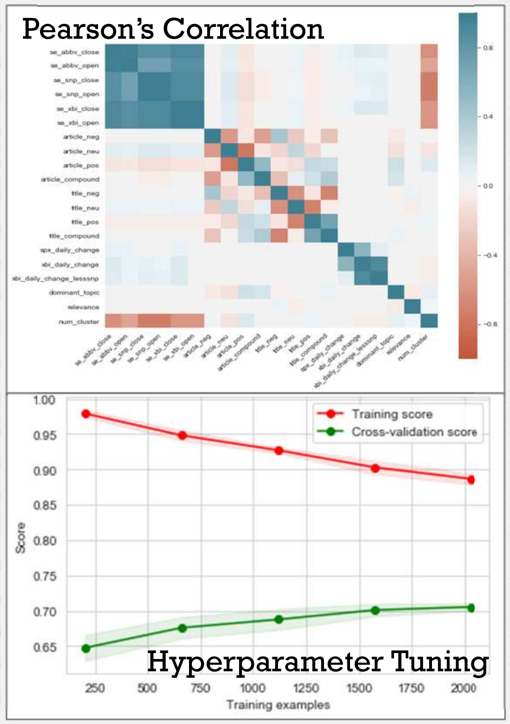


Socio-Economic Risk – Model Findings

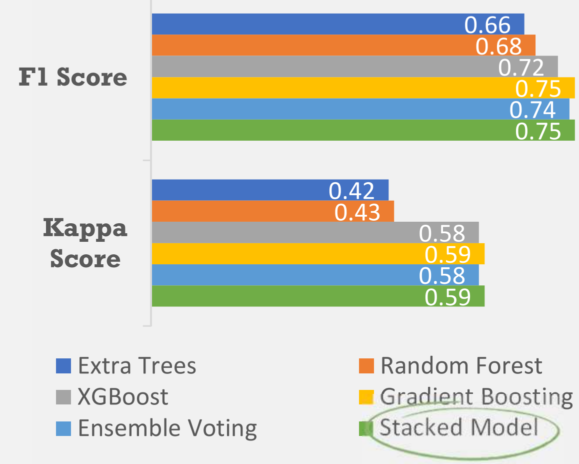
Language Model Findings



Classifier Model Findings

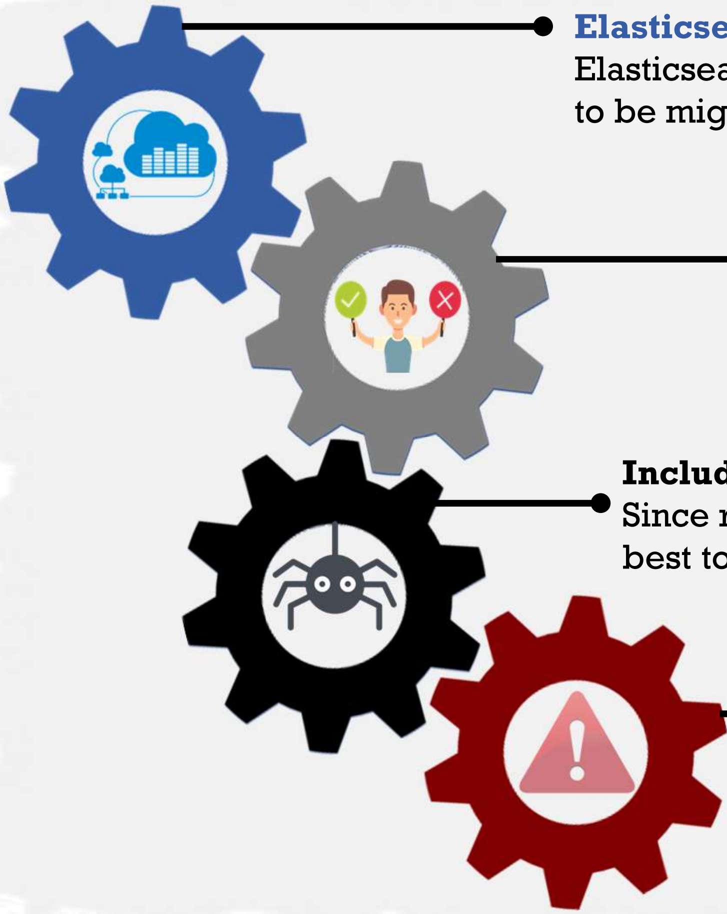


Model Evaluation



Models were evaluated using the Precision, Recall, & F1-Score weighted average metric and the Cohen Kappa Score. After running tuning on all 6 variations, we selected the **Stacked Model** due to its overall evaluation scores.

Recommendations



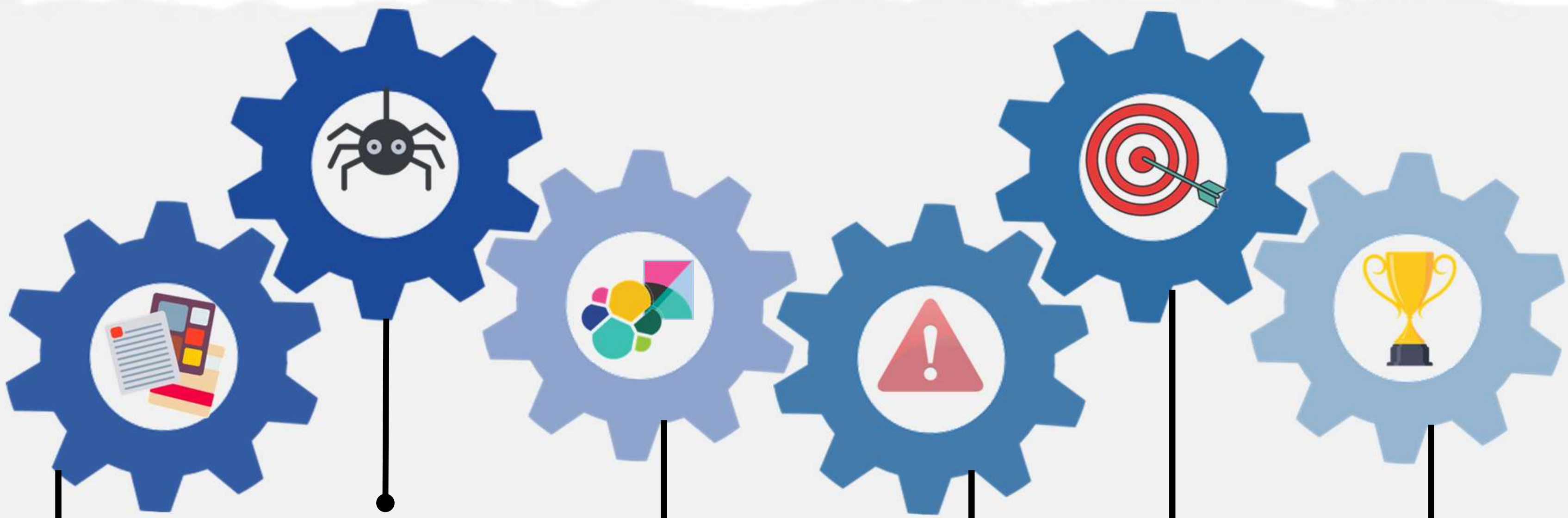
Elasticsearch & AWS migration
Elasticsearch database currently resides in UChicago AWS and needs to be migrated to AbbVie's cloud infrastructure for use

Add feedback loop to user interface
Incorporate active learning in the UI to improve models over time

Include custom crawlers for regulatory websites
Since regulatory websites across countries have different structures it would be best to build custom crawlers for important countries and increase coverage

Review risk labeling
QA team could validate the manually labelled sample to ensure models are built accurately

Conclusion



An invaluable set of **manual article annotations** for risk modeling

A Web Crawler that reviews approximately **10x** as many sources as the manual process

A pilot use case for **Elastic Search** and **Kibana**, which are new to AbbVie

Two new sources of risk data: **Socio-economic** and **Geo-political** risk

Three risk models with **high levels of accuracy**

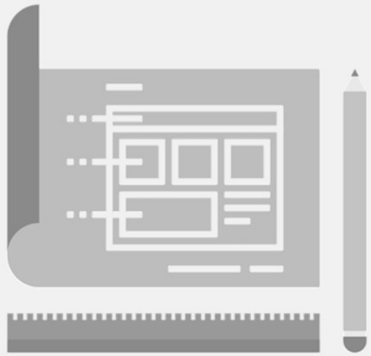
A **successful**, **customizable**, **cost-effective** foundation for end-to-end risk modeling

Agenda

1 Problem Statement

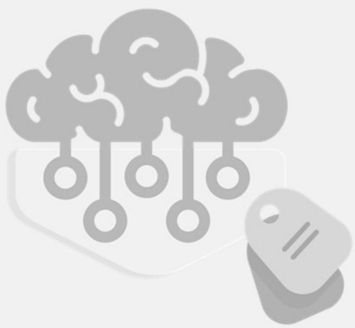
2 Dashboard Demo

3 Architecture



4 Crawler

5 Tagging Models



6 Risk Models

7 Conclusions

8 Appendix

ETL Pipeline

Read repo list of URLs

Crawl latest set of articles from each URL

Language filter

Get stock information

Cleaning meta features such as dates, titles etc.

elastic +

Extract Geospatial information

Extract named entities (such as organizations)

Address missing values

Check for duplicates existing within ES database