Table of Contents

[**Data Collection**](#_7g8lm5xb9h0n) **3**

[Google Search](#_8re0ywkpjaok) 3

[News Articles](#_65u9945nes8u) 4

[Introduction](#_t1ytjxrwju59) 4

[Process Map](#_2s6x30q7wfdq) 5

[GDELT Summary and Feature extraction](#_zifyjhkvv1v4) 5

[Patents](#_t179c3af62nt) 8

[Introduction](#_co7gyv7ilvyg) 8

[Search Engine used: “Google Patents”](#_h4d45hk9j29) 8

[Search string used: "Focal issue + keyword"](#_g15e8vyj9enp) 8

[Patent Scoring:](#_eb8g5b7i2imc) 9

[Social media](#_18hfd4mww56r) 10

[Twitter](#_147iybw736z1) 10

[Tweet Extraction:](#_9qoao58qi219) 10

[Keyword Extraction:](#_lvpcc2624983) 10

[Hashtag Extraction](#_fz264z5jdtzb) 10

[Hashtag Metrics:](#_mn2bf657psya) 10

[Facebook](#_3xz4muandh1q) 11

[Public Feed API:](#_iolrvgo8k0g7) 11

[Graph API:](#_gnkb2wyy8sgp) 11

[VC Firms](#_afvcgve3qp9i) 11

[Introduction:](#_bhfc5dy03lv5) 11

[Data collection on the investments:](#_2lsb5zcla6iw) 12

[Analysis of the Transaction:](#_ktgmtm8c5rxd) 12

[Experiences in the methodologies:](#_9qmwmwxi228y) 14

[Improvising of the signals:](#_u5d49o5df4tl) 15

[Percapita Investment:](#_n63ls3gcbzzq) 15

[Modify the extracted signal:](#_stqs38fmz5j6) 15

[Collecting the description:](#_j4gbk5xqfam7) 15

[**Signal Extraction by DOD and DOV**](#_hvvqd4rspxhp) **16**

[Introduction](#_vhn64sanfsfc) 16

[Collection of Pdfs and Patents](#_v5xsc5if0ki9) 16

[Extraction of text from Pdfs and Patents](#_9i024jtx2utm) 16

[Extraction of abstracts from Pdfs and Patents](#_wkj39gsnhfzg) 17

[Extraction of Phrases from Pdfs and Patents](#_yy8rjbm75w4y) 17

[Filtration of the Keywords](#_adz38kqitcys) 17

[Term Frequency](#_nz1gz1l9q4ti) 17

[Document Frequency](#_ugkjtffnae0b) 17

[Filtering of keywords using Techniques](#_96n1w1phh58t) 18

[Calculation of Degree of Visibility (DoV)](#_arj3xnvbgnoc) 18

[Calculation of Degree of Diffusion(DoD)](#_hvajc5vict6f) 18

[Calculation of Increased rate for both DoV and DoD values](#_j9366572exvk) 19

[Calculation of the average Term Frequency and Document Frequency](#_9kk3bsrxi1sb) 19

[Plotting the Key Issue Map and Key Emergence Map](#_159syfl86n4u) 20

[Strong, Weak, Well Known and Latent Signals](#_jqnm53cgznd0) 21

[Extraction of common signals from DoV and DoD](#_9nwp3enmbj3e) 22

[Extraction of the context of the keywords from the Text corpus](#_h433y85yh3gz) 23

[**Extracting Key Word in Context**](#_me55ztt1a9dh) **23**

[**PESTLE Categorization**](#_y8lcoh3m7o1f) **24**

[**Use of GPT-3 for Scenario generation**](#_8rz61ax9wvgj) **24**

[Users](#_byv72d4953z1) 28

[Process Flow Diagram](#_x12cy4vr5as8) 28

[Login Page](#_ify0smi0duqi) 28

[Client Category Selection](#_vhrfftsrjg1j) 29

[Input Keywords](#_dcuh85b8g5z5) 30

[Signal Generation](#_4zn1bptr3hsa) 30

[Signal Evaluation](#_qrdiydpx8g3m) 31

[Red Team and Blue Team Evaluations](#_lkrx5xrndwz3) 32

[Signal Finalization](#_eh10o7nb6yua) 35

[Scenario Generation and Finalization](#_bcfbknnznnjq) 36

[**Client Category Selection**](#_p5h3b650gzpe) **40**

[**Keyword inputs from Multiverz**](#_7ju2jdv0zyop) **40**

[**Collection of Titles**](#_dtwszsnpedry) **40**

[**Filtration of Titles using PESTEL dictionaries**](#_wyth2u8kxx7j) **41**

[**Titles Vetting by Multiverz**](#_n87n3cyhc05f) **41**

[**Development of signals based on Titles**](#_mcrifxb303im) **41**

[**Collection of signals from GPT-3**](#_4bygkhjz8h8c) **41**

[**Evaluation of Signals by Red Team and Blue Team**](#_9z1mh2ice0lp) **42**

[**Signal Finalization**](#_6xhoxcvmg3s) **42**

[**Development of Scenarios**](#_ubbxn05euznt) **42**

[**Collection of Scenarios from GPT-3**](#_1flktqi4mj7w) **42**

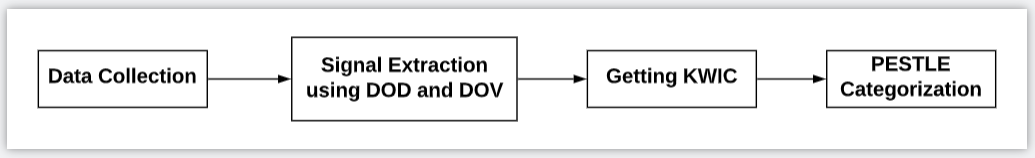
[**Scenario Finalization**](#_qv3c27gk0xkd) **42**

**Introduction**

*Yet to be updated*

Approach 1

In order to identify the signals, data has to be collected and the steps that were implemented initially were discussed in the flow diagram below.



*Fig: Steps implemented in the project*

# Data Collection

The entire process starts with collecting the required data. The resources used are google search, patents, vc firms, news, social media. All these resources are explained in detail.

## Google Search

In this google search process, our aim is to get the most relevant documents related to our focal issue. We need the best search query for getting the relevant results, the search query pattern follows a combination of keywords which are focal issue specific, consulting organizations involved, future related keywords and the timeline which we are interested in. For example we can see in the below sample search query related to Fashion Future

( "Resort collection " OR "Ethical clothing" OR "Fashion show" OR "Sustainable fashion" OR "Buyers")

AND

(government OR World Bank OR IMF OR "World Economic Forum" OR OECD OR McKinsey)

AND

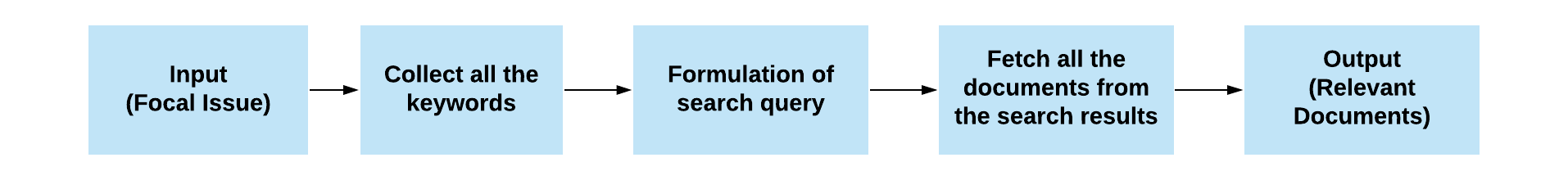
(forecast OR future OR scenarios)

AND

(2030 OR 2040 OR 2050 OR 2060)

**Note**: **AND acts like a connection of two categories**.

**Proposed Model:**

****

Give the Focal issue as input and then collect all the keywords related to the Focal issue either by using the collection of dictionary of keywords related to specific focal issue or by collecting the keywords by using auto generated keywords by using the following methods like Keyword Tool,Uber suggest Tool or Google recommendation. Uber Suggest Tool and Keyword Tool are premium versions. After the collection of the keywords formulate the search query like a modular search query as shown in above example and then we are going to fetch all the documents from the search results.

Note:

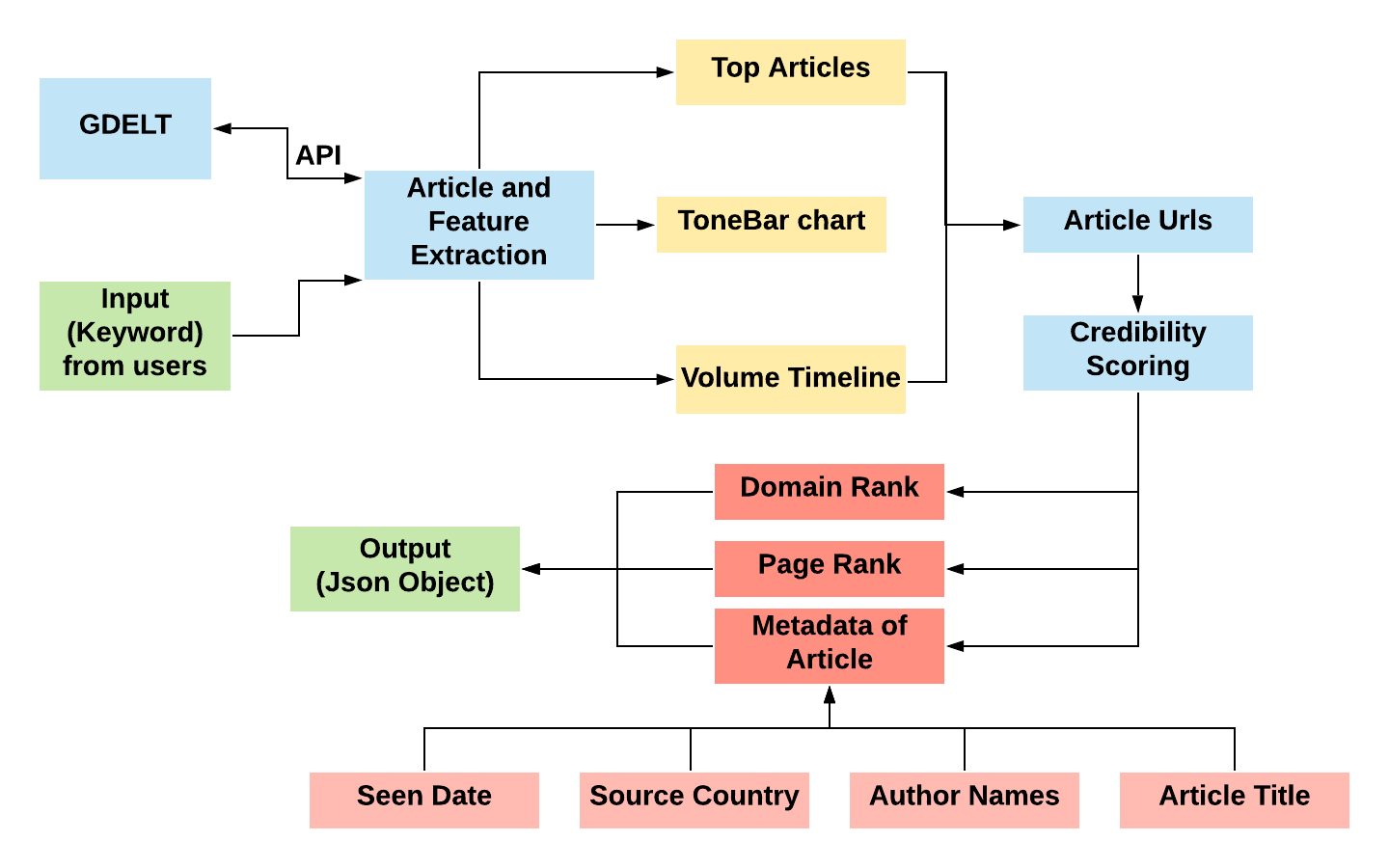
We are getting better results by using Auto generated keywords than a Dictionary of keywords because there are some issues with maintaining a dictionary of keywords like Search Engine constraint, Many search queries, Prioritizing the keywords and Relevancy of the documents.

## News Articles

#### Introduction

In general, people always rely on news to know about events that are happening around them. News acts as a bridge to keep the people updated. Hence news articles are one of the best sources to look for signals. GDELT, a project that deals with the collection of all the news articles, is the world’s largest open-source database about human society. This is a platform that monitors almost everything that happens around the world in print, broadcast, and web formats, in over 100 languages. All the moments were recorded in the form of events which can be referred back whenever required. The oldest moments of this database take you back to 1979. This platform provides an API, which can be used to access the data and use it as and when required.

#### Process Map

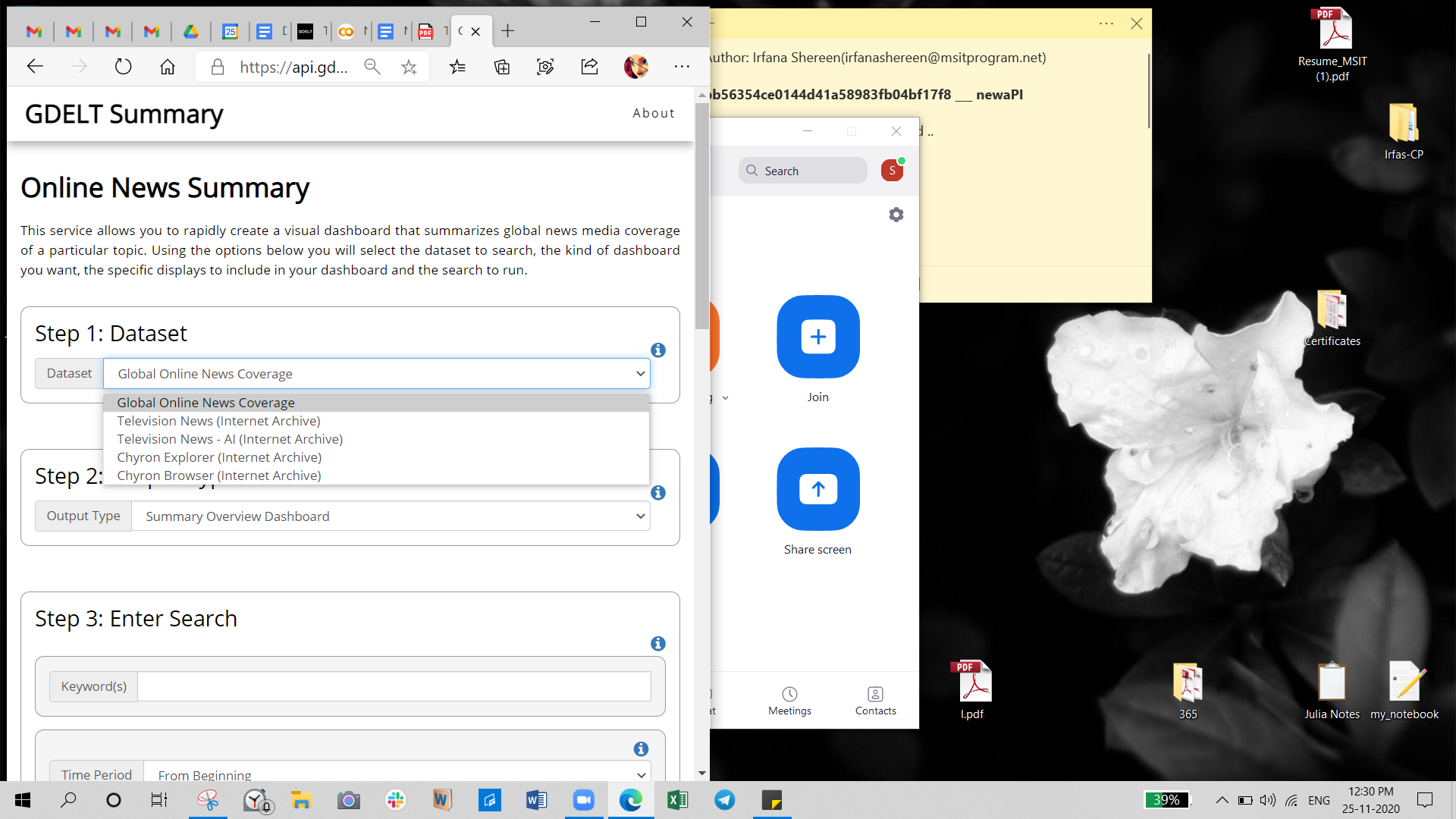


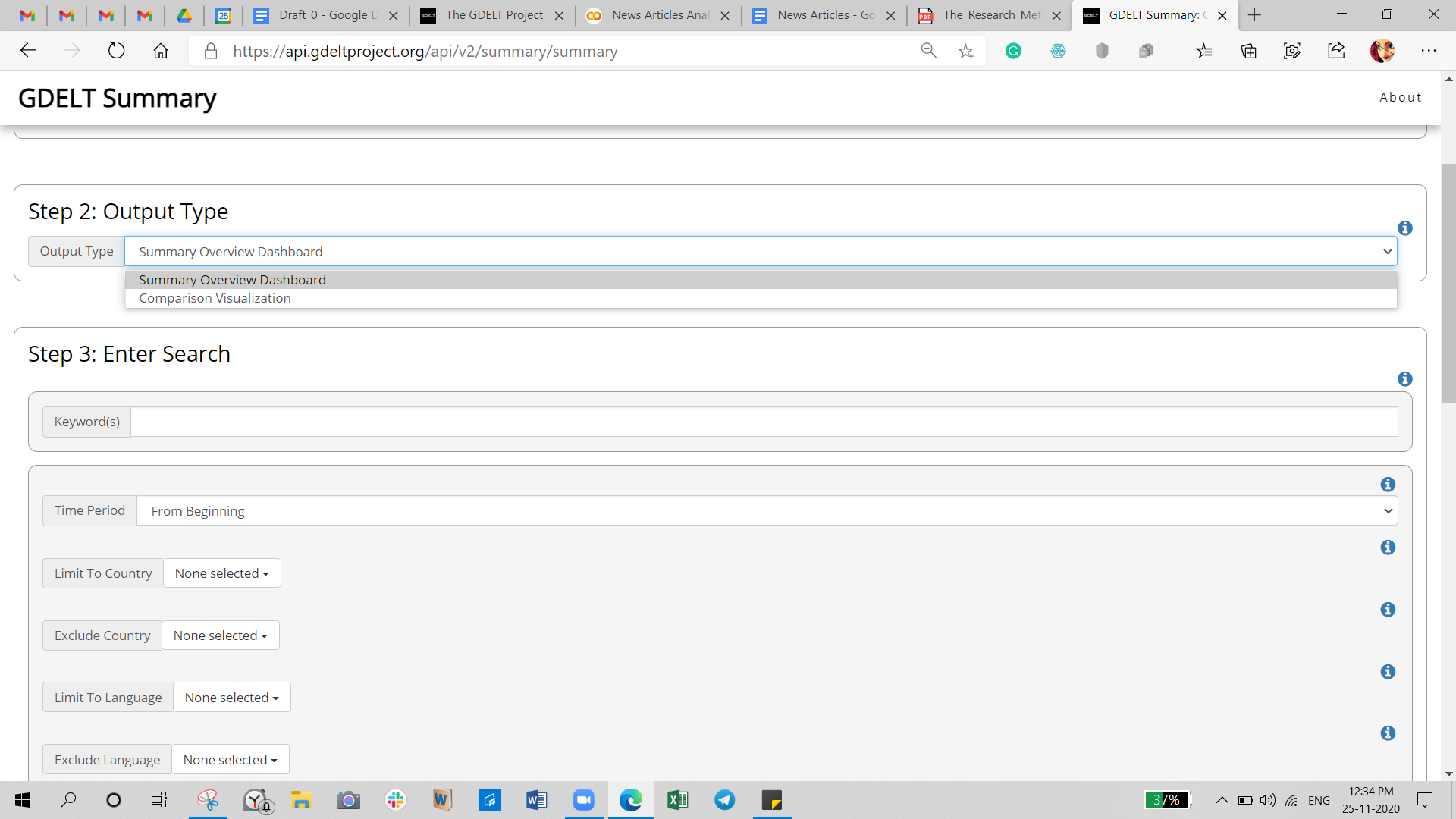
*Fig 1: Representation of all the processes involved in the NEWS article module*

The above figure depicts the complete process that is being followed in this module. It starts with taking the input from the user in the form of a keyword and the output will be a JSON object which is used in the further processes. In detailed explanation of all these steps are given below

#### GDELT Summary and Feature extraction

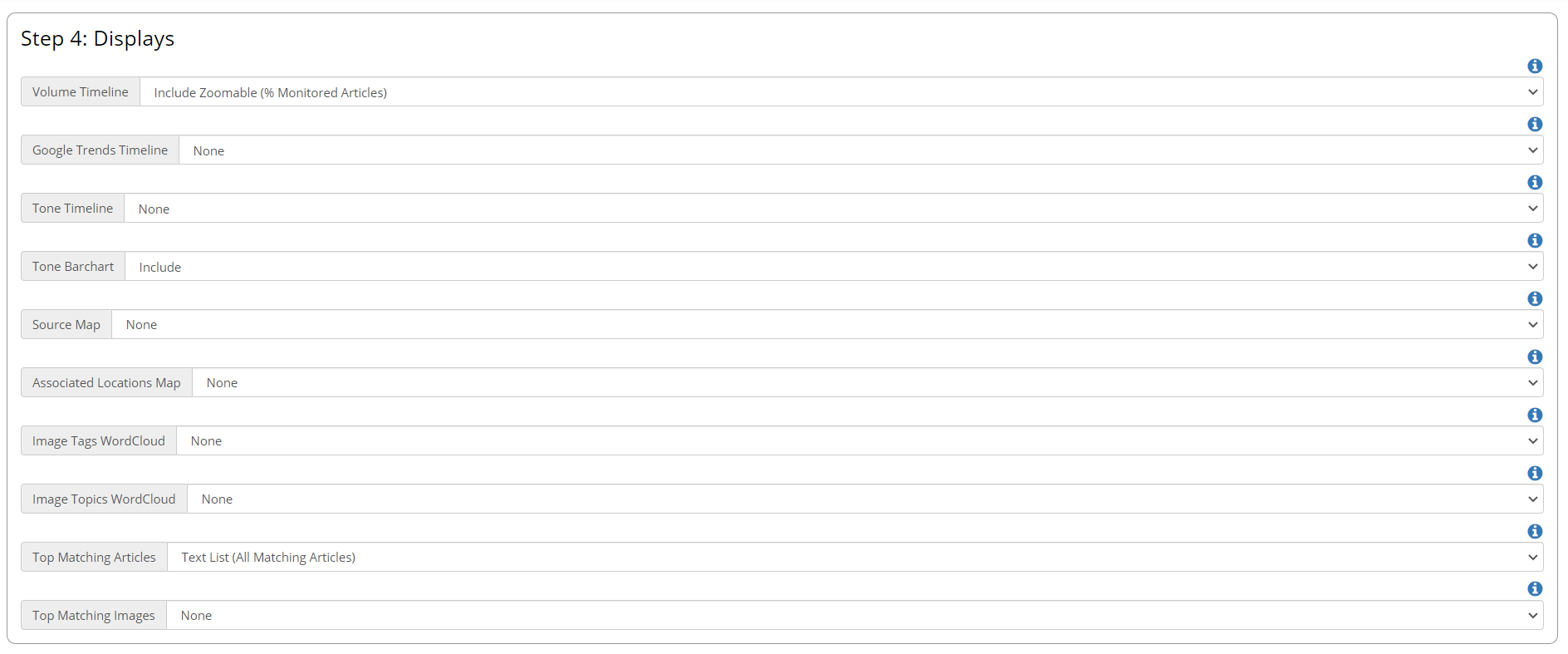
This is one of the services that GDELT offers and it provides a custom dashboard with inbuilt analytics. It takes inputs from the users in 4 steps, Dataset,Output Type, Enter Search, Displays. All the options that were there in these steps are shown in the figures (2-4) below.

*Fig 2: Representation of all the processes involved in the NEWS article module*

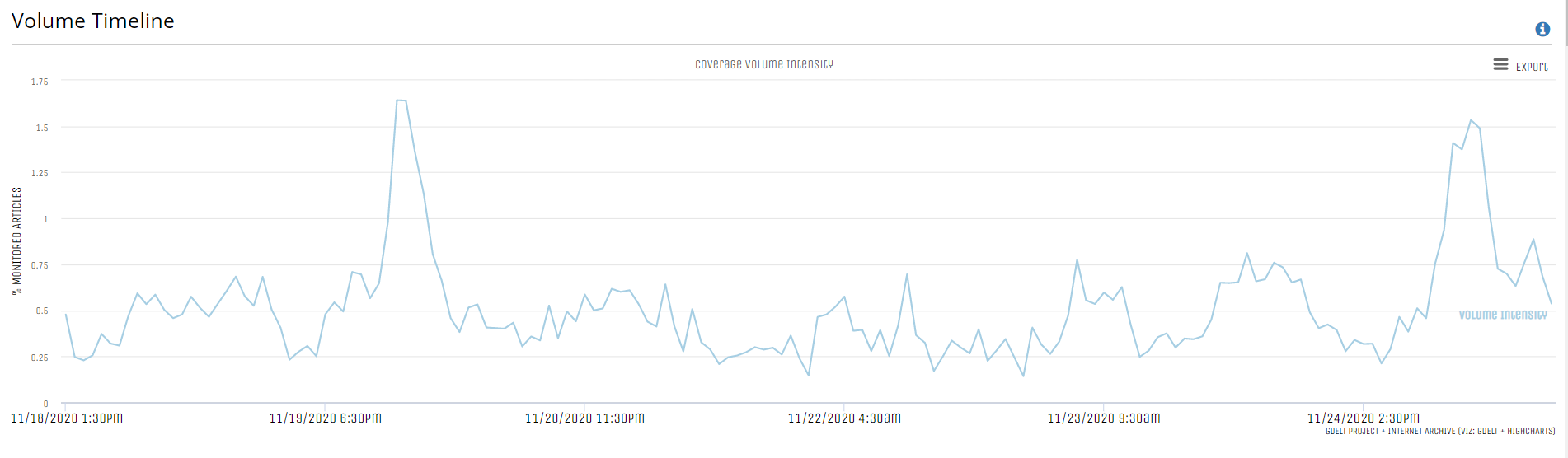


*Fig 3: Representation of all the processes involved in the NEWS article module*

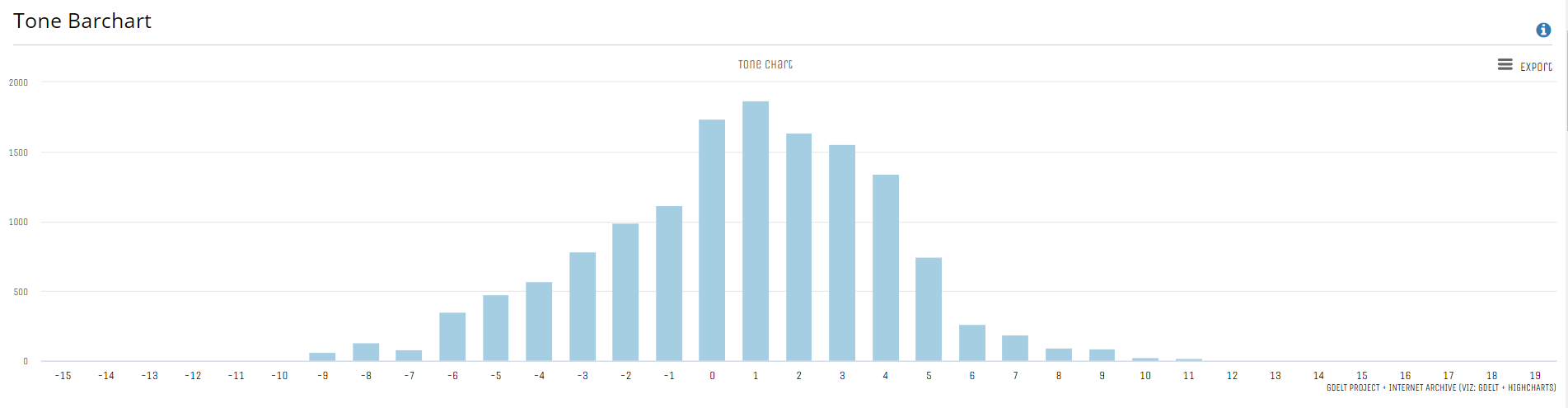


*Fig 4: Step3 & Step4 of the input collection* 

Here for the keyword “Fashion”, the following are the features that we are interested in. and the results can be seen in the figures. The line graph in figure 5 shows the percent of all global online news coverage monitored by GDELT over the given day/hour that matched your query, allowing you to trace how attention to your topic has changed over time and whether it is increasing or decreasing.



*Fig 5:Volume TImeline Representation*



*Fig 6: Tonebar chart representation*

The tone-barchart above(Fig:6) is a unique display that takes all matching coverage monitored by GDELT over the selected time period, computes the document-level tone of each article (ie, the tone of the article as a whole rather than the tone just referring to your keywords) and bins them into a histogram/barchart that ranges from extremely negative on the left to extremely positive on the right. By hovering over each bar, you can see a few of the topmost relevant matching articles that were scored with that tone level.

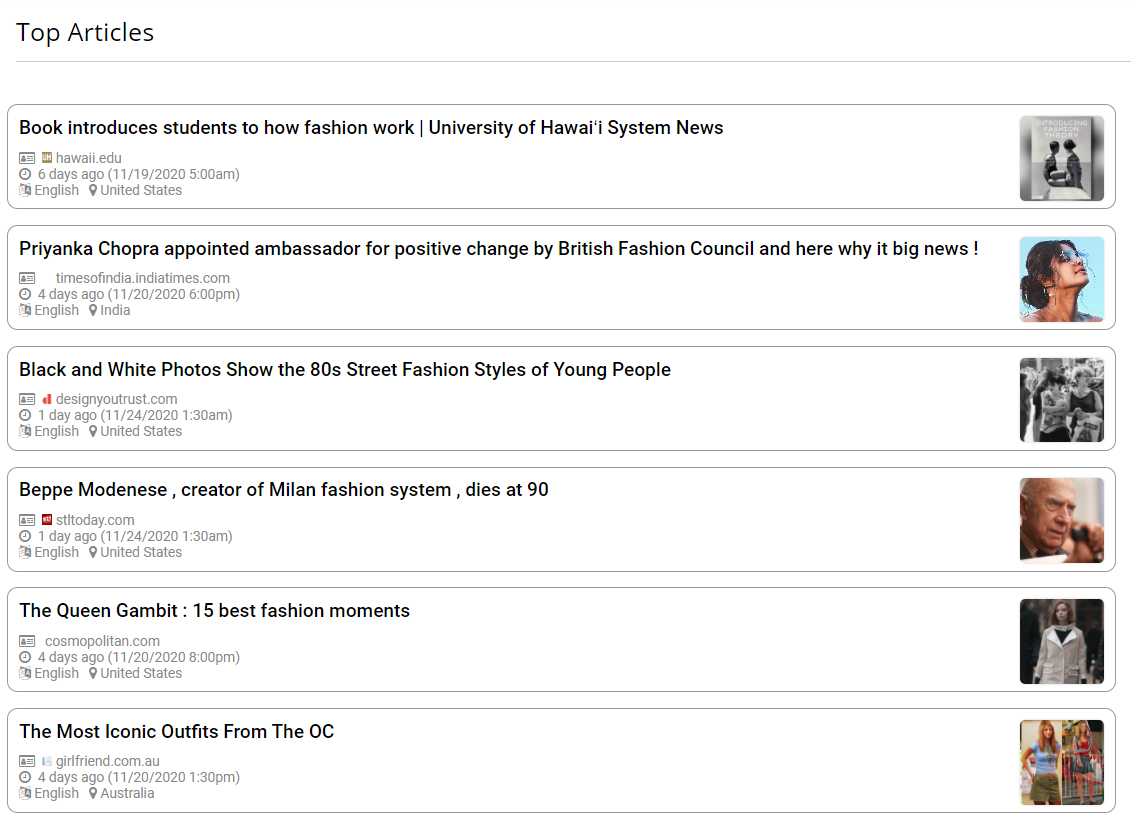
*Fig 7:The top articles for the given keyword*

Figure 7 represents a short list of some of the most relevant coverage that matched your search. The number of articles depends upon the topic/keyword given by the user. If the topic of interest is very niche, there will not be many articles and if the topic of interest is very common, then there will be many articles, here for “Fashion”, 33 different articles were displayed.

All these data can be exported in many different types like Pdf, JSON, CSV. With the use of API, you can download the data automatically to your program file(here Google Colab notebook). For each feature which was mentioned earlier, there are separate API calls and the data can be stored as a JSON object for further analysis. The sample API request and the response for the Volume timeline are given below.

**API Request for Volume timeline**

[*https://api.gdeltproject.org/api/v2/doc/doc?format=json&timespan=1W&query=Fashion sourcelang:eng&mode=timelinevol&timezoom=yes*](https://api.gdeltproject.org/api/v2/doc/doc?format=json&timespan=1W&query=Fashion%20sourcelang:eng&mode=timelinevol&timezoom=yes)

**Response Object(truncated)**

{"query\_details": {"title": "Fashion sourcelang:eng", "date\_resolution": "hour"}, "timeline": [ { "series": "Volume Intensity", "data": [ { "date": "20201118T080000Z", "value": 0.4773 },..........]}]}

## Patents

#### Introduction

Patent documents provide a rich source of technological and scientific knowledge that can reveal technological trends as well as information on the landscape of the market. This makes analysis of the vast and ever-growing number of patents an important part of the scenario creation process. Patents also play a significant role in understanding a company's IP portfolio.

#### Search Engine used: “Google Patents”

Google Patents search engine is the most popular open source patent search engine available on the market which provides access to complete text of patents . It includes over 120 million patent publications from 100+ patent offices around the world, as well as many more technical documents and books indexed in Google Scholar and Google Books, and documents from the Prior Art Archive. The sheer number of patents available on the search engine and the fact that it is an open source tool makes it the ideal choice to use it as our search engine to find patents.

The search can also be narrowed down by only including specific fields within your search scope. We have taken use of one of these features to extract year wise patents for the dod and dov analysis by modifying the priority data field. Google patent search allows users to keep their search restricted to only major fields like Title, Abstract, and Claims.

#### Search string used: "Focal issue + keyword"

The keywords used are the most prominent terms extracted for each focal issue from the subject experts. For example, for fashion we have taken keywords from the website ShilpaAhuja.com which is one of India’s most-read digital fashion magazines.

Example search string: "Fashion Knitwear"

#### Patent Scoring:

Each search operation has produced over 100000 results, the patents produced at the tail end of the search results drifted away from the focal issue and if these patents are included in the corpus created the results from the DOV DOD analysis might be tainted.

Even though every patent has its own importance, some patents have higher value as they have a bigger geographical presence and have acted as a base point from which further inventions are created.

We have used 3 types of scoring methods which will help us define the worth of patents from various dimensions. The scoring mechanism is designed after going through various papers on quantitative scoring methods for patents and also trying to replicate the methodologies of some of the most popular and paid patent scoring tools like AcclaimIP.

**Citation Score:**

Analysing patent citation is an important technique which can be used to study the evaluation of science and technology. The basic principle behind its use is that patents cited as prior art by many later patents tend to contain important ideas upon which numerous inventions have been built.

Citation scoring uses a number of forward citations as the metric for scoring, so let us understand what it actually means.

Forward citations - Citations marked Forward are patents that came after the document you are viewing. Forward references are patents or applications that cite or reference the patent in question.

When the forward citation count is more that means the current patent has played a very prominent role in further developments in that respective field, so higher the citation score higher the value of the patent.

**Technology Score:**

Technology score is the metric that indicates how technologically diverse a patent is.

Every patent is classified by human curators during the time of filing by the Wpo. The system used to classify patents is the The Cooperative Patent Classification, it is in force from 1 January 2013.

It is a bilateral system which has been jointly developed by the EPO and the USPTO. It combines the best classification practices of the two offices.

Their classification consists of nine sections, A-H and Y, which in turn are sub-divided into classes, sub-classes, groups and sub-groups. There are approximately 250 000 classification entries.

**Geographical Score:**

This score provides us with an understanding of the geographical presence of the patents.If a particular patent is filed in multiple countries then it indicates that the patent has more value.

## Social media

In 2020, around the world almost 3.6 billion people engage in these platforms through which individuals, communities, and organizations can share, co-create, discuss, participate and modify content or self-curated content posted online.

### Twitter

Twitter is one of the social media platforms where the users can tweet (post) and retweet (share content). Other people can retweet it and share the post among their followers. We will use this platform to extract the signals. In order to find the signals we will be using tweet extraction and hashtag extraction.

#### Tweet Extraction:

Using the query given by the user a Search query is created(For fashion Example : ‘ “fashion” AND “futures” ’). Using the twitter API and search query, we were able to extract the past 7 days( since running the algorithm) tweets from twitter.

### 

#### Keyword Extraction:

The tweet/retweet consists of the following:

1. Text msg()
2. Hashtags (Optional, organisations, plain text, other users, twitter groups etc..)
3. Urls (Optional, website/video/audio etc.)

Using the website urls from the tweets we will extract the page content and combine it to the text msg of the tweet. Using NLP(Natural Language Processing) to clean the text (Removing the multiple spaces, and invalid characters(!@#%^).Remove stop words (the,that,and) ). After filtering the text we are using [Monkey Learn API](https://monkeylearn.com/text-extractors/)(keyword Extractor) to get the keywords from the text.

#### Hashtag Extraction

Using the twitter API and search query we were able to extract the past 7 days( since running the algorithm) hashtags from twitter.

#### Hashtag Metrics:

Since it will be too difficult to identify the impact of a hashtag, Hashtag metrics are introduced here. 4 metrics are being calculated based on various parameters.

1. **Volume of mentions**: It’s an elementary hashtag metric. It’s the number of pieces of content mentioning your hashtag.Tracking this metric gives you an idea of how popular the hashtags are.
2. **Volume of interactions**: Social media platforms such as Facebook, Instagram, or Twitter give users the chance to interact with them: like, comment, share or retweet a piece of content. Those are called social media interactions. By following this metric, you can assess how engaging the social media posts with hashtags are. The more interactions the better.
3. **The most influential authors**: THis is about the authors mentioning your hashtag.It’s an insight into social media profiles mentioning your hashtag of a significant social media influence. Social media influence is calculated based on the number of followers, activity and the number of interactions of a source. It’s based on an assumption that influential sources have a huge audience and hashtag will be exposed to a wide range of audience.
4. **The most interactive authors**: These are the hashtag mentions from sources with the largest number of interactions. This is how you can identify sources who are most interactive and drive discussions around anything.

### Facebook

There are two types of API’s provided by facebook for extracting facebook data.

#### Public Feed API:

The Public Feed API provides a feed of user status updates and page status updates as they are posted to Facebook.Only status updates that have their privacy set to ‘public’ are included in the feed.

The feed only includes basic data about the given post. From that basic data you may use the Graph API to request additional metadata to supplement the updates received through the public feed API.

Access to the Public Feed API is restricted to a limited set of media publishers and usage requires prior approval by Facebook. The application process to access the API has been terminated and no new applications are being accepted by the facebook team.

### 

#### Graph API:

The Graph API is the primary way to get data into and out of the Facebook platform. It's an HTTP-based API that apps can use to programmatically query data, post new stories, manage ads, upload photos, and perform a wide variety of other tasks.

Graph API Components:

1. Nodes
2. Edges
3. Fields

Graph API can be used to access public groups and public pages data but the data to public feed has been restricted.

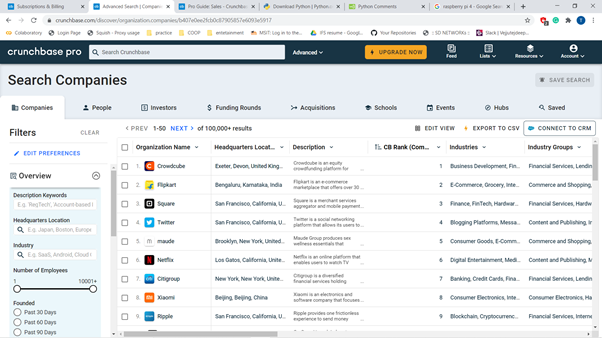
## VC Firms

#### Introduction:

Our aim is to isolate the investments more than $150 million to analyse those investments and the field they are invested in. These investments are taken as signals for the development of the sector in the future. In order to do those, we have approached many methodologies and few worked out and many didn't. I would like to discuss in the following section what experience gained through each methodology and how I analysed the investments.

#### Data collection on the investments:

The investment data are collected from the Crunchbase website (<https://www.crunchbase.com/home>). This website contains a database on the VC investments that are done all over the world. To get those data we have to subscribe to the pro version of the website which might cost around $29 per month. Then those data can be viewed on the website and they can be downloaded into the local system using the option called CSV format. On the website it shows around 10,000 transactions and we can only export 1000 transactions at a time. The database is updated on a daily basis and new transactions appear on the website daily. We like to select the column to investments, IPO and number of the investors in the VC using the option edit view.

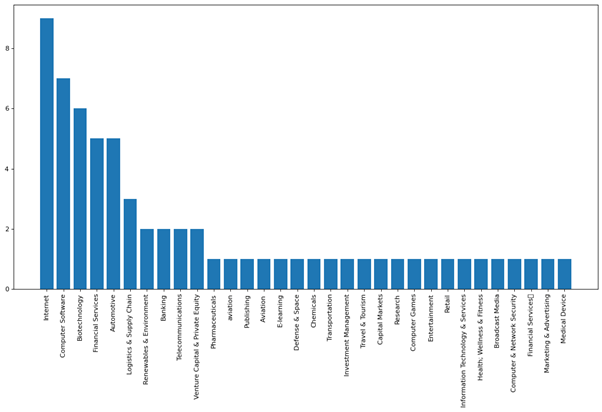


*Fig 8: Extraction of data from crunchbase*

#### Analysis of the Transaction:

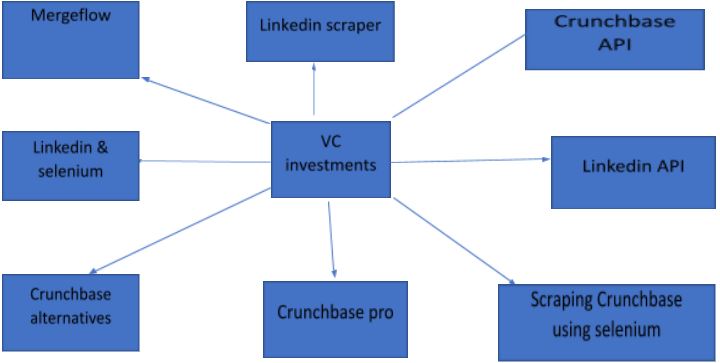
Now we have the 1000 VC investment data on the local website. We isolate the columns required for the analysis using pandas.

Then we group columns by the money in the investment in the venture capital and try to sum the investments in that and visualize it to get how the investments are distributed over the different sectors. How each sector grows in the perspective of the future. As shown in the below picture the internet (online shopping) has been receiving major capital and y axis is the investments in the billions taking the sector as whole we can deduce that the internet (online shopping) would be a major industry in the future and play a major role in the fashion sector. The IPO would look into the growth of the company and look into how much the company had growth since the establishment of the industry. We would like to know how much each investor has invested into the industry.



*Fig 9: plotting of the results from the extracted data*

#### Experiences in the methodologies:



*Fig 10: pursed methods*

Firstly, we tried to access the Crunchbase API but couldn’t get the access and after connecting to the Crunchbase team they said it would cost around $8000 at least for the access of the CrunchBase API

Next, we identified the alternatives for the CrunchBase and came up with the few of them and tried to get the required data from the site called Mergeflow other alternatives were unreliable or just too costly to use,

Initially Mergeflow seems a reliable site to use, but the further we research into the site the further we find the loop holes in the data after researching for one week we find out that the site was too unreliable to use.

Next, we tried scraping from CrunchBase through the selenium and found out that the Crunchbase has a strong protection called distil network. They are one of the best bot detection services out there. So that was of no use in going through that approach.

Next, we try to get half of the data from the CrunchBase and the remaining we try to get from Linkedin. From Linkedin we were trying to find which industry does the organization belong to in the sector. For example Marrow optics company belongs to the optical industry in the fashion sector.

We tried different methodologies to access the Linkedin that are.Firstly, we try to us the Linkedin API for extraction of the data to access the API we need to set an app and that need the permission from the legit company to access the data of the organization and we should apply for the Linkedin organization request, and otherwise we can only get the data from our connection. Afterwards we need to follow a complex process to get the data from Linkedin. Which was too complex to automate so we couldn’t use the Linkedin API.

After that we tried scraping Linkedin using selenium. This process was working fine for 5 requests. But we have huge corpses of data which are too filled from Linkedin. So, we tried more than 5 requests, after 5 requests the Linkedin server stops responding to the requests and then sends text that “bots can’t be used to scrape this site.”. I tried to increase the time gap between the requests but it didn’t work.

Finally, we tried to use the Linkedin crawler to extract the data. It was working fine for the person's profile but it didn’t work for the organization profile; it just responded with the empty text for the request from the crawler for the organization.

After that we received information related to the industry in the sector from the CrunchBase pro itself the only disadvantage was to upload a CSV file from the Crunchbase pro to the UI we are developing in order to get the emerging technologies.

#### Improvising of the signals:

#### Percapita Investment:

Firstly, we try to consider only the investment made in the sector for considering it as emerging technology. It needs to have a minimum investment of the $150millions. Now we are considering the number of investors in the sector and taking the quotient of it with the investment made in the sector. We do it to take the order of presidencies in the technologies we consider emerging.



Percapita investment = 

#### Modify the extracted signal:

We use the condition vc investment more than 150 million to isolate the required data from the given dataset. If the output data is less than 10 percent of the whole data for example if there are 500 rows if the output data is less than 50 rows then the data is not sufficient for the prediction. That is why we included the condition that the output data is less than 10% of the whole data then instead of the condition of companies greater than 150 million investment we include the top 10% of the whole dataset. So we could get the required data.

#### Collecting the description:

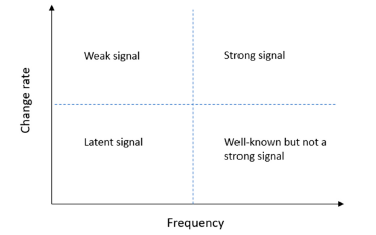
When we collect the technology there is a single word and send the data to scenario we couldn’t create a good scenario from the data available. So we collected the description of the industry and sent that to the tool to create the quality scenario for that industry.

# 

# Signal Extraction by DOD and DOV

## Introduction

The aim of this methodology is to identify the future signals of the given focal issue by using text mining tools. These signals are used for identifying the drivers of change, which plays a crucial role in generation of the scenarios. Initially we have to curate the content from the Articles and Patents from different sources. After extracting the text from the documents we have to apply text analytics using DoV and DoD. By using the term frequency, document frequency , DoV, DoD values, we have categorized the filtered keywords into four quadrants. If the signal has a high frequency and change rate then it is a Strong signal. If the signal has a high frequency and low change rate then it is Well-known but not a strong signal. If the signal has a high change rate but low frequency then it is Weak signal. If the signal has a low frequency and change rate then it is a Latent signal.



Signal Categorization

## Collection of Pdfs and Patents

We have collected Pdfs which are relevant to focal issues by using the google search from Articles, journals and stored in the repository. We have collected and stored a required number of Pdfs per each year from 2010 to 2020 in their respective period wise repositories. For patents we have collected the required number of patents from the google patents each year and gave priority to the top ranked patents which are based on the patent scoring process.

## Extraction of text from Pdfs and Patents

We have extracted the text from the Pdfs by using the python pdf scraping module like PyPDF2 and for patents we have used beautiful soup for extracting text like description.

## Extraction of abstracts from Pdfs and Patents

We have extracted abstracts from the Pdfs and if there was no abstract then we have extracted conclusion and summary. After extracting we have stored these abstract data separately for each file in respective periods further, which was used for filtering the key words criteria. For abstract extraction also we have used python packages for both pdfs and patents.

## Extraction of Phrases from Pdfs and Patents

Abstraction of phrases was used to extract the phrases in the text without losing the importance of the phrases. For abstraction phrases we have used the Gensim python package(bigram/ trigram). We have followed the same process for both pdfs and patents in extraction of phrases from Pdfs and Patents.

## Filtration of the Keywords

We have used the python NLTK package for tokenizing the sentences and converting to each word. After converting to individual words we have filtered keywords by applying stop words using NLTK and Spacy python packages. Stop words means grammatical words like articles, punctuations, conjunctions.. Which are not necessary for our methodology where we can reduce noise in the text corpus.

## Term Frequency

After filtering the words we have to find the term frequency for each word in all the documents in each period(year). Term Frequency means the frequency of each word in that period to the total number of unique words of all the documents in that period.

TF = Frequency of each word in all documents of that period **/** Total number of unique words of all the documents in that period.

## Document Frequency

we have to find the document frequency for each word in all the documents in each period(year). Document Frequency means the number of documents where the keyword occurs to the total number of documents in that period.

DF = The number of documents where keyword occurs **/** The total number of documents in that period.

## Filtering of keywords using Techniques

We have filter the keywords by using the NLP techniques :

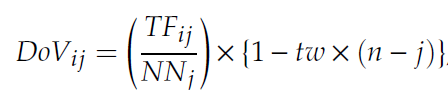
* Checking the keyword is valid english word or not by using NLP synsets package
* Removing Stop words, trailing spaces, unwanted special characters, punctuations
* Removing adverbs using parts of speech tagging
* The keyword must occur at least ten times in the text corpus of each period(year)
* The keyword must occur in at least five document abstracts.

By using these techniques we are able to cut the noise in the text corpus and filter the keywords to the large extent.

#### 

## Calculation of Degree of Visibility (DoV)

DoV measures the degree of the frequency of a defined keyword in a set of documents is presented by a proxy for signal. The signal dimension of a future sign is related to the visibility of the future signal. To study the visibility of a sign, the absolute frequency of occurrence of each word in a set of documents(for every year) is measured. This is defined as the Degree of visibility of the keyboard i in period j (specific year eg: 2010) as can be seen in the below equation.

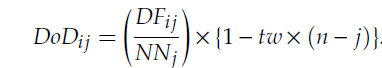


* Where TFij is the total number of occurrence of the word i in period j(considering all the documents in each period)
* NNj is the total number of documents in period j(period)
* n is the number of periods
* tw is a time weight and the tw has been defined as 0.05 by a group of experts.

We have calculated the DoV value of each keyword by using the above equation.

## Calculation of Degree of Diffusion(DoD)

Degree of Diffusion measures the document frequency of each keyword in relation to the total number of documents as a proxy for issue. Both DoD and DoV indicators put more weight on recent occurrences through a time-weight coefficient.



* Where DFij is the number of documents in which the word i appears in period j.
* NNj is the total number of documents in period j(period)
* n is the number of periods
* tw is a time weight and the tw has been defined as 0.05 by a group of experts.

We have calculated the DoD value of each keyword by using the above equation.

#### 

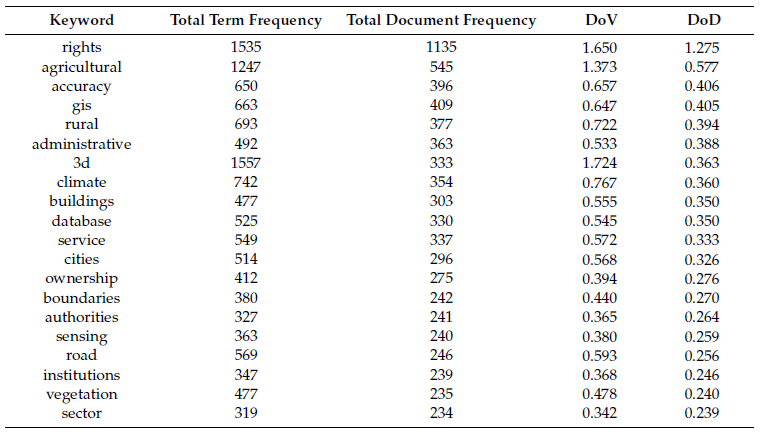
## Calculation of Increased rate for both DoV and DoD values

Increased rate means it is the difference of DoV value of successive periods for the common words between the two periods. Similarly we are finding increased rates for DoD values also. We have calculated the increased rate for both DoV and DoD of every keyword per each year.

## Calculation of the average Term Frequency and Document Frequency

Average Term frequency means it is the sum of the document frequency of the common word between the two successive periods to the total duration of the periods.

Ex: DoD and DOV for Land Administration shown in below table



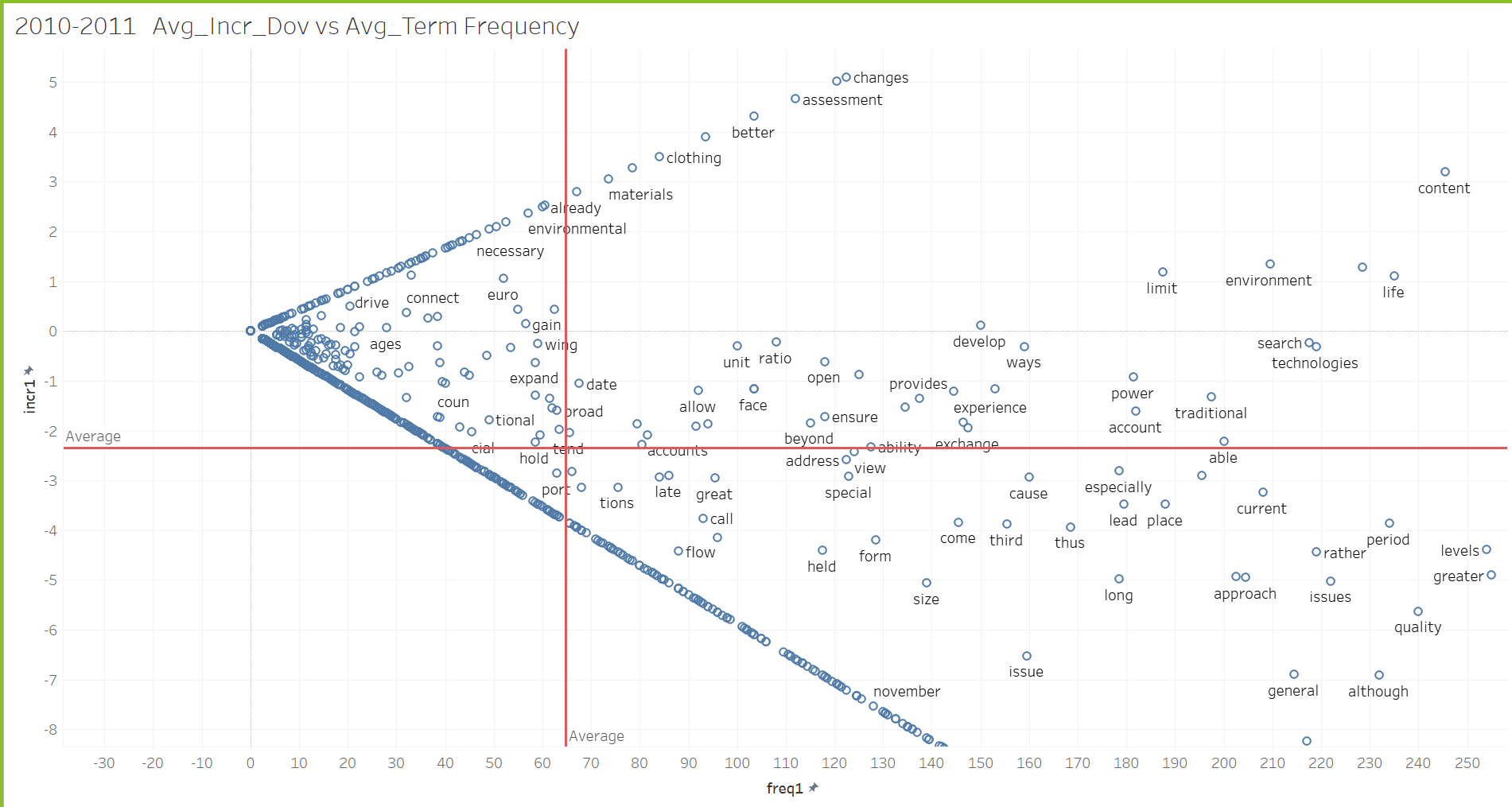
## Plotting the Key Issue Map and Key Emergence Map

Key Issue Map is the plot between the average document frequency on x axis and Increase rate of DoD values on y axis. We have to take the mean value of average document frequency and draw the vertical line on the x axis using the mean average document frequency. Draw the horizontal line with the mean increase rate of DoD on the y axis.

Key Emergence Map is the plot between the average Term frequency on x axis and Increase rate of DoV values on y axis. We have to take the mean value of average term frequency and draw the vertical line on the x axis using the mean average document frequency. Draw the horizontal line with the mean increase rate of DoV on the y axis.

From the above process we can categorize the signals into four quadrants.

**Average Increment rate of DoV vs Average Term Frequency**



**Key Emergence Map**

**Average Increment rate of DoD vs Average Document Frequency**

#### 

**Key Issue Map**

#### Strong, Weak, Well Known and Latent Signals

#### 

* Strong signals : High average term/document frequency and high increase rate of DoV/DoD means keywords with term/document frequency greater than mean term/document frequency and change rate of DoV/DoD greater than mean change rate of DoV/DoD.
* Weak signals : Low average term/document frequency and high increase rate of DoV/DoD means keywords with term/document frequency less than mean term/document frequency and change rate of DoV/DoD greater than mean change rate of DoV/DoD.
* Well-Known signals : High average term/document frequency and low increase rate of DoV/DoD means keywords with term/document frequency greater than mean term/document frequency and change rate of DoV/DoD lower than mean change rate of DoV/DoD.
* Latent : Low average term/document frequency and low increase rate of DoV/DoD means keywords with term/document frequency lower than mean term/document frequency and change rate of DoV/DoD lower than mean change rate of DoV/DoD.

Strong: [ 'invention', 'type', 'power', 'light', 'photovoltaic' ,'surface', 'device', 'sunlight', 'system', 'temperature', 'solar','energy', 'generating', 'efficiency', 'high', ’'water', 'present', 'connected', 'utilization', 'collector', 'conversion', 'angle', 'collection', 'model', 'battery', 'storage', 'electricity', 'cost', 'film', 'lens', 'side', 'reflection', 'module', 'tube', 'mirror', 'thermal', 'unit', 'plate', 'panel', 'radiation', 'axis', 'concentration', 'tracking', 'focus', 'control', 'tower', 'supply', 'heating', 'structure', 'wind' ]

Weak: [ 'field', 'technical', 'including', 'turbine', 'plurality', 'advantages', 'bottom', 'collection', 'according', 'guide', 'belongs', 'curved', 'focused', 'absorption', 'photoelectric', 'plate', 'present\_invention', 'reduces', 'steam', 'construction', 'installed', 'plant', 'discharge', 'production', 'structure', 'integrated', 'direct', 'upper', 'output', 'incident', 'method', 'heating', 'improves', 'effect', 'panel', 'focal', 'formed', 'cells', 'ratio', 'mirror', 'pipe', 'supply', 'pump', 'working\_medium', 'range', 'rate', 'focus', 'kind', 'line', 'located', 'spectrum', 'spotlight', 'storage', 'tower', 'utility\_model', 'wind','silicon', 'layer', 'service', 'power\_generation', 'equipment', 'condensation', 'electricity', 'tube', 'reflection', 'thermal', 'battery', 'present', 'unit', 'control', 'cost', 'surface', 'radiation', 'temperature', 'film', 'axis', 'side', 'module', 'concentration', 'comprises', 'collector', 'angle' ]

Latent: [ 'large', 'ratio', 'pump', 'working\_medium', 'receiving', 'location', 'spectrum', 'spotlight', 'utility\_model', 'service', 'power\_generation', 'equipment', 'condensation', 'effect','focal', 'cells', 'whole', 'multiple', 'guide', 'belongs', 'curved', 'photoelectric', 'control', 'present\_invention', 'apparatus', 'construction', 'installed', 'plant', 'discharge', 'production', 'integrated', 'comprehensive', 'direct', 'upper', 'output', 'silicon', 'layer', 'incident', 'method', 'provides', 'improves', 'realizes', 'cost', 'plurality', 'advantages', 'bottom', 'medium', 'pipe', 'radiation', 'axis', 'concentration', 'supply', 'field', 'technical', 'main', 'focus', 'fixed', 'focused', 'unit', 'absorption', 'first', 'tower', 'wind', 'panel', 'film', 'turbine', 'lens', 'side', 'tracking', 'relates', 'module', 'tube', 'steam', 'structure', 'heating', 'using', 'angle', 'plate', 'collector', 'reflection', 'mirror', 'battery', 'kind', 'storage', 'electricity', 'collection', 'comprises', 'conversion', 'utilization', 'temperature', 'generating', 'thermal' ]

Well known: [ 'thermal', 'efficiency', 'water', 'solar', 'energy', 'high', 'system', 'model', 'device', 'conversion', 'generating', 'kind', 'photovoltaic', 'sunlight', 'light', 'surface', 'arranged', 'power', 'connected', 'utilization', 'battery', 'present', 'mirror','secondary', 'temperature', 'collector', 'lens', 'reflection', 'invention', 'type', 'module', 'angle', 'electricity', 'storage', 'medium', 'pipe', 'comprises', 'radiation', 'film', 'turbine', 'axis', 'side', 'concentration', 'tracking', 'tube', 'focus', 'unit', 'control', 'focused', 'first', 'tower', 'using', 'collection', 'plate', 'supply', 'absorption', 'heating', 'structure', 'wind' ]

## Extraction of common signals from DoV and DoD

After finding the four types of signals for each year in the repository for both DoV and DoD. Then we have to take common signals from both DoV and DoD analysis. It means we have to take common strong signals from DoD and DoV, common Latent signals from DoV and DoD, common weak signals from DoD and DoV and common well known signals from DoD and DoV.

## Extraction of the context of the keywords from the Text corpus

We have extracted the context of the above signals from the text corpus in order to know more information about the signals whether it is impactable or not.

Finally we are able to find the Strong signals, Latent signals, Weak signals and Well known signals from this methodology. After this we need Human Intervention for picking the important future signals by considering the context of the signals generated.

# **Extracting Key Word in Context**

Keyword in Context (KWIC) is extracted because only presenting keywords makes little sense to the reader and also while providing the input for scenario creation.

So to get the context of where a certain keyword has occurred we take the help of KWIC.

In traditional KWIC there is a parameter called window size which decides the number of words present on either side of the keyword but we are considering the whole sentence and not explicitly sticking with any window size.

Example: If “shopping” is the keyword generated then just by looking at the keyword we might not be able to makeout why this keyword is important to the scenario we are building, but by providing the left and right context of the keyword we can analyze if the keyword is important in the context. We can see the KWIC example below.

|  |  |  |
| --- | --- | --- |
| **Left Context** | **Keyword** | **Right context** |
| Poshmark’s mission is to make | **shopping** | simple and fun by connecting people around a shared love of fashion, while empowering entrepreneurs to become the next generation of retailers", |
| com, an interest news site, Eater, food and dining publication, Racked, a | **shopping** | beauty, and fashion publication, Curbed, a real estate and home brand publication, and Re/code, a tech business publication", |
| We are also the only European player to have recently entered the top 25 | **shopping** | apps by daily active users", |
| Lazada is the number one online | **shopping** | and selling destination in Southeast Asia \\xe2\\x80\\x93 present in Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam", |
| Working closely with its partners, Global Fashion Group has crafted in-class | **shopping** | experience for its customers offering numerous international and local brands to be delivered quickly and conveniently", |
| Verishop aims to be a go-to lifestyle | **shopping** | destination whenever the customer in search of elevated brands and quality-driven products", |

# PESTLE Categorization

PESTLE (Political, economic, social, technological, legal and environmental) analysis is a framework that helps to identify all of the various external factors that might affect a business.

The PESTLE categorization is being done so that the keywords and sentences extracted from the dod/dov analysis and GDELT process are sorted into meaningful categories, this will help us in fine tuning the input given for scenario creation by varying the percent of data taken from each category.

With the help of PESTLE dictionaries which contain all the keywords related to that category, we have sorted the extracted keywords and sentences based on the word count.

Example: “Indian law makers and representatives have long been the pawns of industrialists as they fund their election campaigns”

The above sentence has “law”, “election”. “Campaign” , three words from political dictionary so this sentence falls under the political category.

# Use of GPT-3 for Scenario generation

**Introduction:**

GPT-3 is a pre-trained transformer model which has been trained using 175 billion parameters. It can be used for a variety of applications like Text generation,classification,transformation,text completion, conversations etc.

The idea behind using this model in our project was to generate text for scenarios using the keywords,sentences extracted from the text corpus or even giving incomplete scenarios as input.

There is one advantage of using GPT-3 to other text generation models like BERT is that we need not provide any training data to generate text, and also because it has been trained and tested using different learning mechanisms.

1. **Few-shot learning(or in-context learning) :** where we allow as many demonstrations as will fit into the model’s context window (typically 10 to 100),
2. **One-shot learning:** where we allow only one demonstration
3. **Zero-shot learning:** where no demonstrations are allowed and only an instruction in natural language is given to the model.

One shot and Zero shot learning are the difference makers as the model has been trained to perform with minimal to zero training data and automatically fine tune for a new task.

**Core Components:**

1. **Prompt -** The input given to the GPT-3 API to generate completions for, encoded as a string, a list of strings, or a list of token lists.
2. **Completion -** The text generated by the API when a prompt is given.
3. **Tokens-** The API turns text into tokens before processing it. One token is roughly 4 characters for normal English text.

Example : “Pear” - ‘Pear’ , “Descartes” - ‘Desc’ + ‘art’ + ‘es’

The Tokenization method used here is Byte pair encoding (bpe). Both the Prompt and Completion are limited to 2048 tokens or roughly 1500 words

**Engines:**

Engines are the background models that have been trained. GPT-3 has 4 such engines that have been pre trained so that they excel at different tasks and also some engines that are good at everything.

1. **Davinci:**

Davinci is the most capable engine and can perform any task the other models can perform and often with less instruction.The trade-off with Davinci is that it costs more to use per API call and other engines are faster.

**Good at:** Complex intent, cause and effect, summarization for audience

1. **Curie:**

Curie is quite capable for many nuanced tasks like sentiment classification and summarization. Curie is also quite good at answering questions and performing Q&A and as a general service chatbot.

**Good at:** Language translation, complex classification, text sentiment, summarization.

3) **Babbage:**

Babbage can perform straightforward tasks like simple classification. It’s also quite capable when it comes to Semantic Search ranking how well documents match up with search queries.

**Good at:** Moderate classification, semantic search classification

4) **Ada:**

Ada is usually the fastest model and can perform tasks like parsing text, address correction and certain kinds of classification tasks that don’t require too much nuance. Ada’s performance can often be improved by providing more context.

**Good at:** Parsing text, simple classification, address correction, keywords

**Parameters:**

1. **Sampling:**

Two Parameters in GPT-3 for sampling:

* 1. **Temperature sampling**:( Range varies from 0 to 1)
     1. It is inspired by statistical thermodynamics. It controls the randomness of the model.
     2. Lower temperatures make the model increasingly confident in its top choices.
     3. 0 temperature is equivalent to max likelihood.
     4. Higher temperature values means the model will take more risks. For example 0.9 gives more creative answers and is used in creative application.
  2. **Top-p sampling**: ( Range varies from 0 to 1)
     1. The model considers the results of the tokens with top\_p probability mass. Controls the diversity of models.
     2. So 0.1 means only the tokens comprising the top 10% probability mass are considered.

1. **Max\_tokens (Integer)** : Requests can use up to 2048 tokens shared between prompt and completion
2. **N (Integer)**: How many completions to generate for each prompt.
3. **Echo (Boolean)** : Echo back prompt along with completion.
4. **Presence\_penalty (0 - 1)**: Penalizes new tokens based on whether they appear in the text so far. Increases the model's likelihood to talk about new topics.
5. **Frequency\_penalty (0-1)**: Penalizes new tokens based on their existing frequency in the text so far. Decreases the model's likelihood to repeat the same line verbatim.
6. **Best\_of (Integer)**: generates multiple(based on the integer value) completions server side and gives only the best. Every completion is charged.
7. **Logit\_bias**: Accepts a json object that maps tokens (specified by their token ID in the GPT tokenizer) to an associated bias value from -100 to 100.

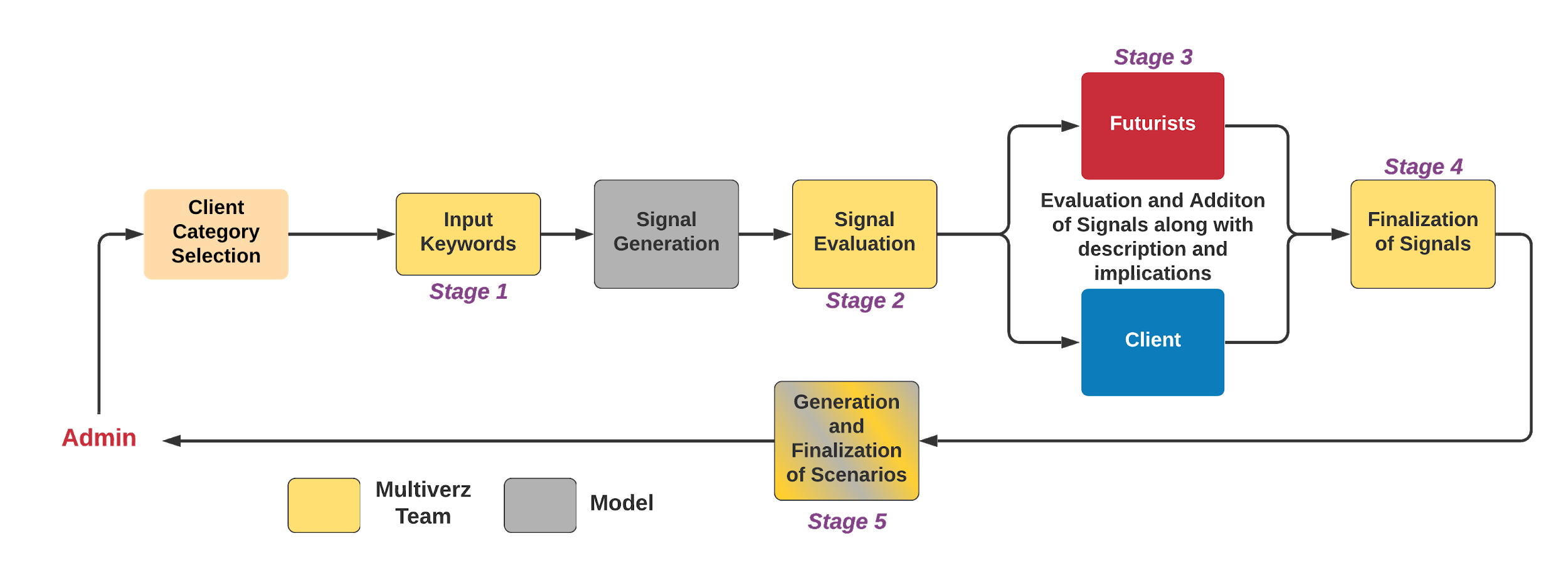
Prototype Design

## Users

There will be 4 different user accounts for this application. Three of these users will be working in the backend by providing inputs and validating the results at multiple levels.

1. **Admin**: The one who raises a request initially and the scenarios which will be generated in the last stage are shared to the admin. The Admin account is like an end user/customer who makes use of this application.
2. **Multiverz Team**: This is one of the backend teams, which provides and initiates the model by giving all the necessary inputs. This team is also responsible for finalizing the signals and scenarios.
3. **Futurist**: One one the backend teams, who will rate the signals and provide the justification for the same.
4. **Client**: part of the backend team to evaluate and rate the signals by providing the justification just like the futurist.

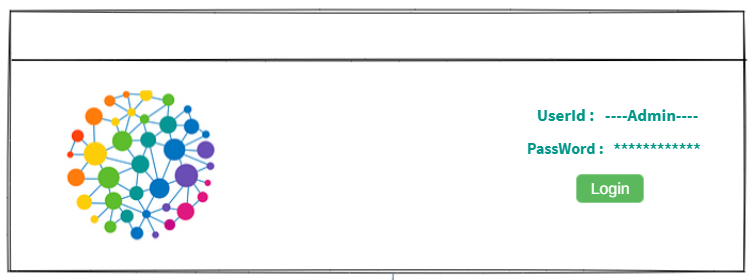
## Process Flow Diagram

The flow chart(refer to Fig 10) shows the entire pipeline of the project. It has 5 stages and each stage has its own value add in terms of generating effective output. These steps are discussed below along with the UI Screen designs. 

*Fig 10 : Entire Process Pipeline of generating scenarios*

### Login Page

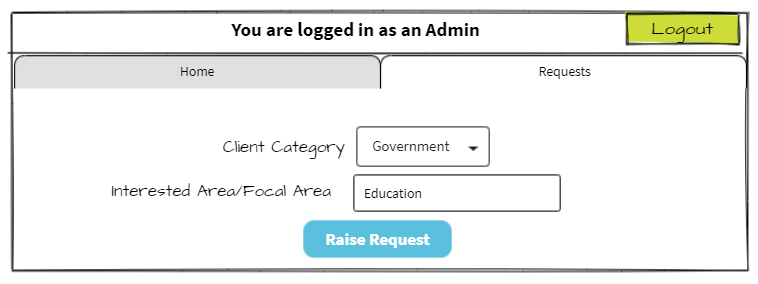
Any user who wants to access the application needs credentials(refer Fig 11). All the four user accounts will have different credentials, which allows you to access the corresponding pages.



*Fig 11 : Login Page*

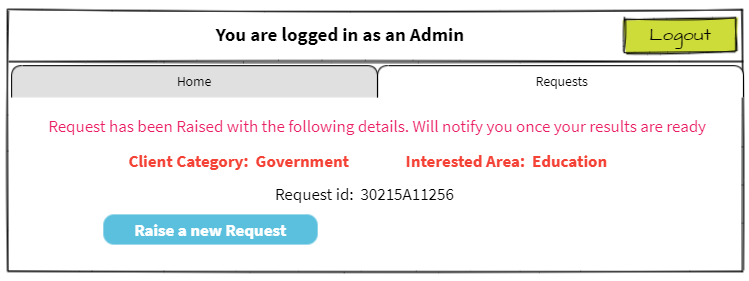
### Client Category Selection

This is the first step in the process of generating the scenarios(Refer Fig 12). The details of the scenarios which you want to generate will be given by the Admin. There are 3 different clients for this application - Government, Organization and Educational Institutions. The Admin needs to select one from these categories. In addition to this, the Admin also gives the focal area for which the scenarios should be generated.



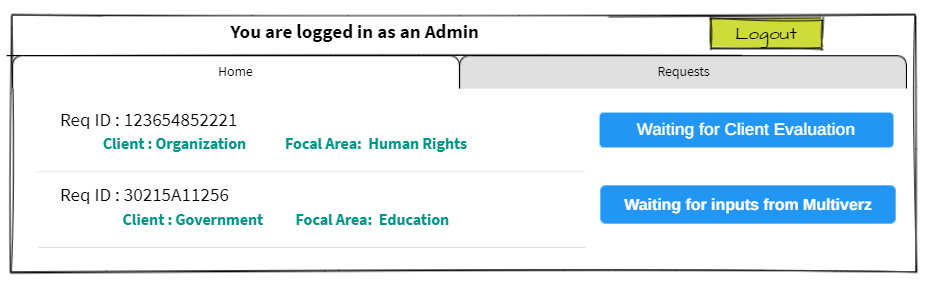
*Fig 12: Category selection by the admin(by Admin )*

Once the admin raises a request, a notification will be shown to the admin with all the details and a request id(Refer Fig 13). The admin can raise as many requests as needed.



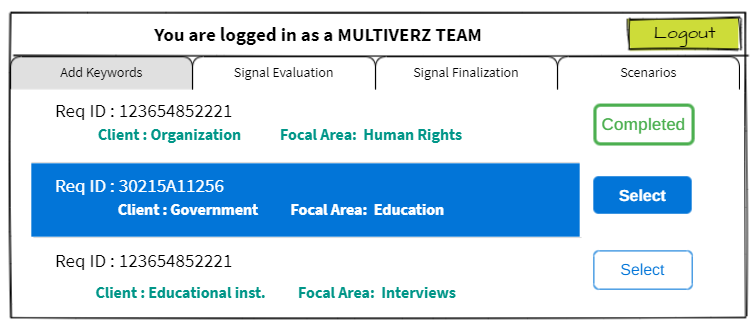
*Fig 13 : Request Initiation(by Admin )*

All the requests raised by the admin will be shown in the Requests tab(Refer Fig 14) along with the stage at which the request is. This will be similar to that of a tracking request.



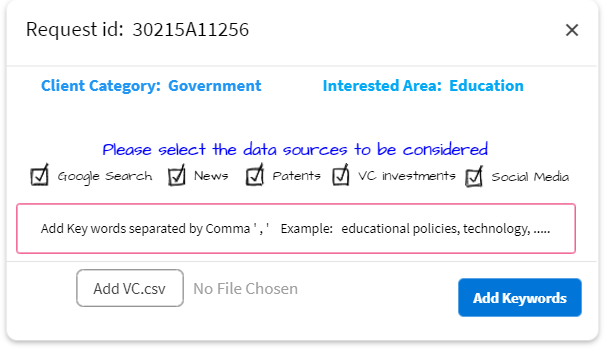
*Fig 14 : Status of all the requests initiated by Admin*

### Input Keywords

Once the request has been raised by the admin, the request will be forwarded to the Multiverz team, who will start the initial process of adding all the necessary keywords. There will be a notifications page, where the team can see all the requests raised by the admin(Refer Fig 15)

*Fig 15 : Requests/notifications Screen for adding Keywords (by Multiverz )*

Select the request id for which you want to provide the Keywords and the screen shown will appear(Refer Fig 16), where the team can select the sources to include and upload the required files.



*Fig 16 : Input Keywords for the Request raised by Admin (by Multiverz )*

Once the inputs are given by the team, they will click on “Add Keywords” and then that request will be shown as completed in the notification page.

### Signal Generation

After all the required inputs from the Multiverz team, the model starts collecting the signals in the background and will notify the user when they are ready.

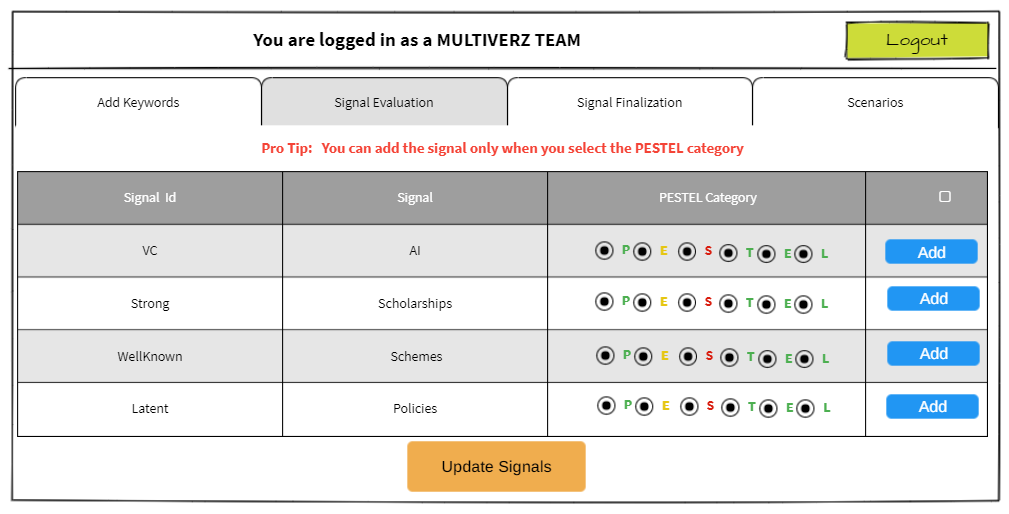
### Signal Evaluation

All the signals that were generated by the model were shown to the Multiverz team as in the first level of filtration. All these requests will be displayed in the Signal Evaluation tab(Refer Fig 17).



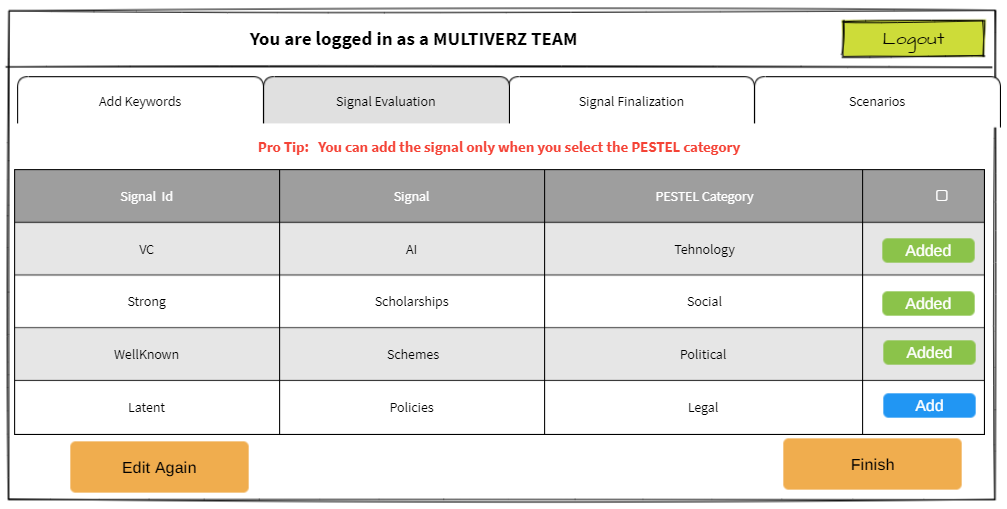
*Fig 17 : Requests/notifications Screen for Signal Evaluation (by Multiverz )*

Click on the request which you want to evaluate, then all the signals along with their respective sources are displayed. The team needs to select the signals which they think will be necessary and categorize them under PESTEL(Refer Fig 18)

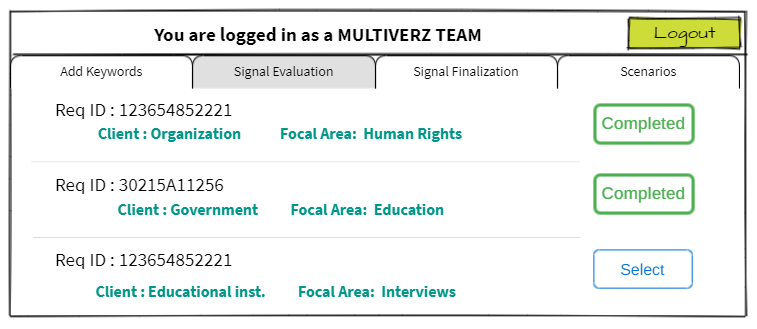


*Fig 18 : Stage 1 of Signal Evaluation(by Multiverz )*

Clicking on Update Signals, will show a confirmation page(Refer Fig 19), where all the selected signals will be shown. If you are not satisfied with the signals, you can click on the Edit Again option, which will take you to the previous page((Refer Fig 18). If you are ok with the signals and want to confirm, click on Finish. The request will be shown as completed in the notifications page(Refer Fig 20).



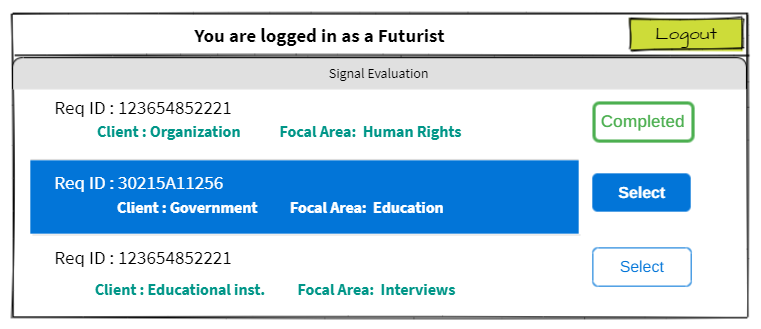
*Fig 19 : Finalized Signals(by Multiverz )*



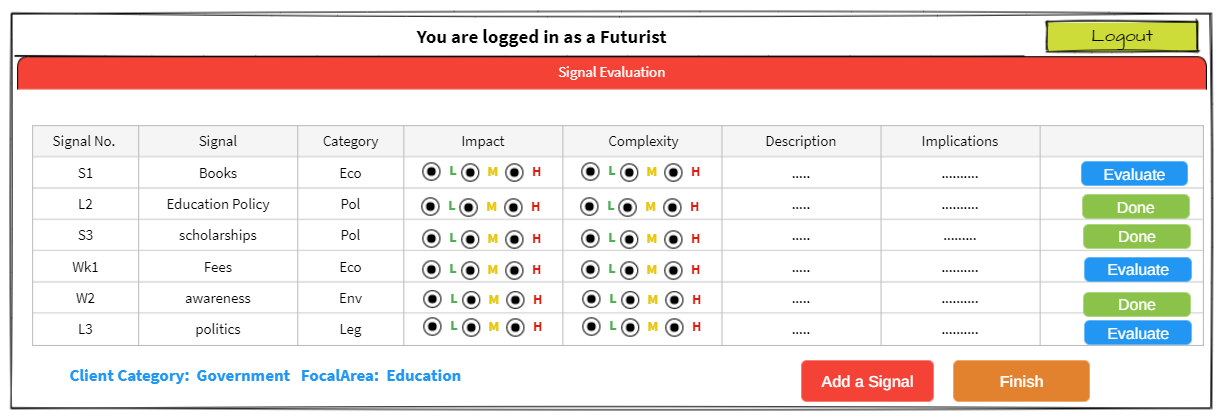
*Fig 20 : Signal evaluation completed status(by Multiverz )*

### Red Team and Blue Team Evaluations

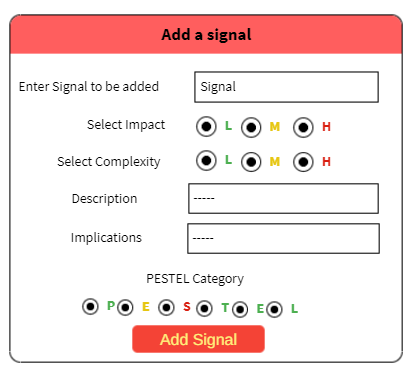
After the first stage of signal selection from the Multiverz team, these signals will be forwarded to Futurists and Clients, simply known as Red Team and Blue Team respectively. In this stage both the red teams and blue teams will be able to rate the signals and write the description, implications corresponding to each signal(Refer Fig 22 and 25). If the team wants to add a new signal they can add it at this stage, simply by clicking on the Add signal button and a new pop up comes in(Refer Fig 23 and 27). They too have a similar requests page just like the others(Refer Fig 21 and 25). After finishing the evaluation, the status of the request will turn to completed(Refer Fig 24 and 28)



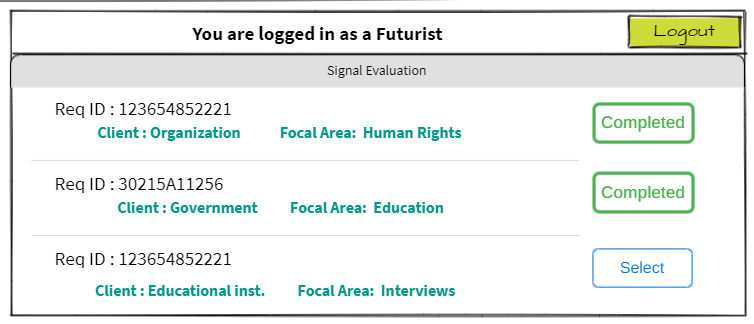
*Fig 21 : Requests/notifications Screen for Signal Evaluation (by Futurist )*



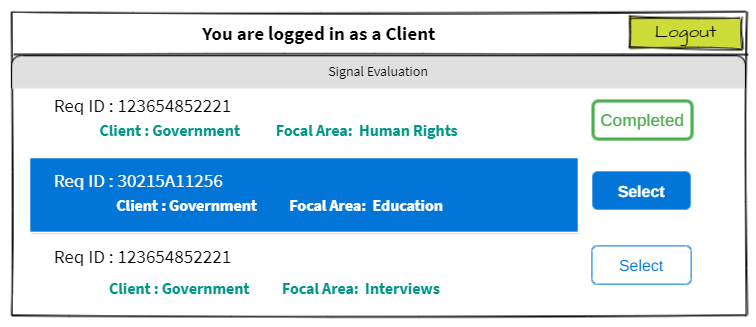
*Fig 22 : Signal Evaluation by Futurist*



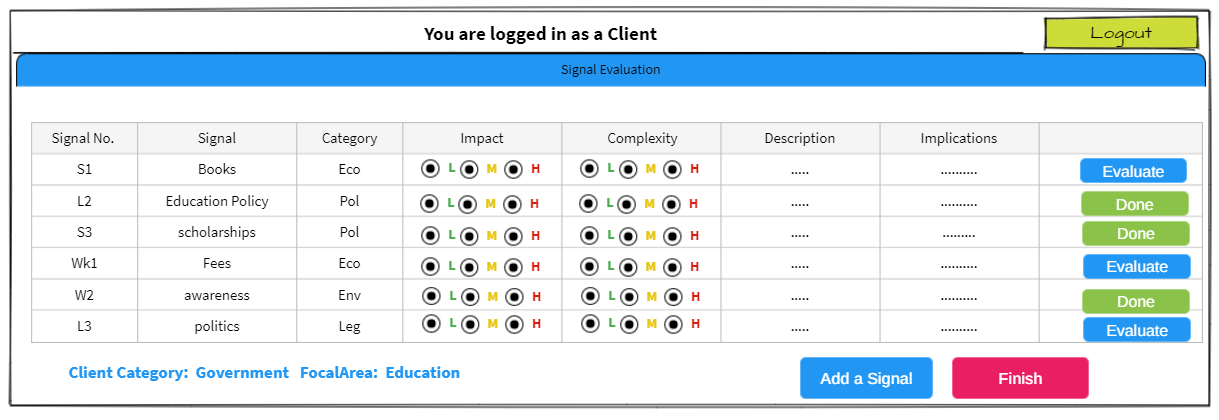
*Fig 23 : Signal addition by a Futurist*



*Fig 24 : Signal evaluation completed status(by Futurist )*



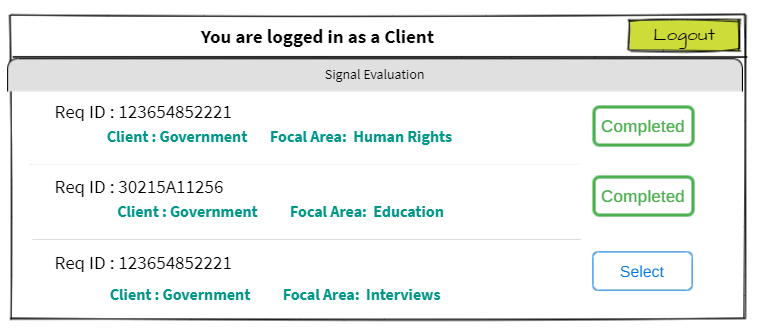
*Fig 25 :Requests/notifications Screen for Signal Evaluation (by Client Screen)*



*Fig 26 : Signal Evaluation by client*



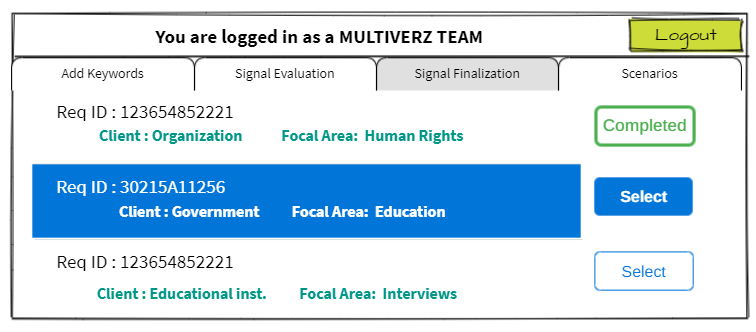
*Fig 27 : Signal addition by client*



*Fig 28 : Signal evaluation completed status(by Client)*

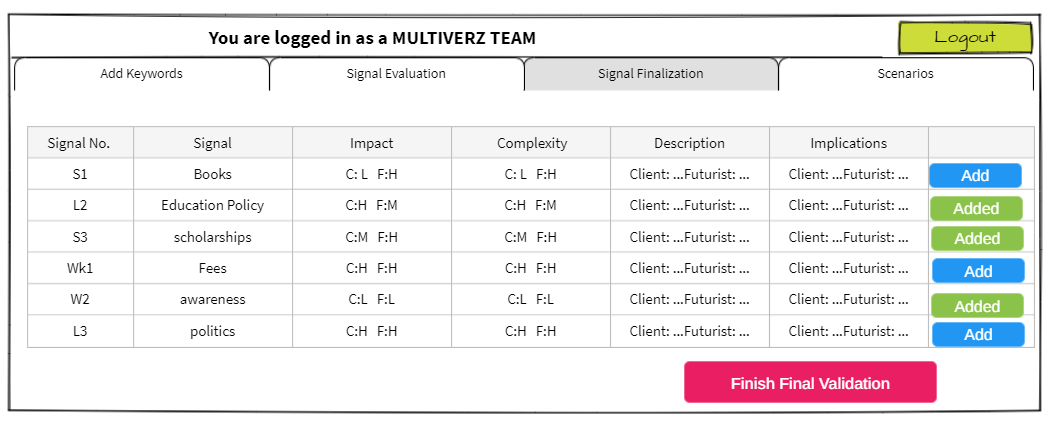
### Signal Finalization

All the signals along with the implications and description are forwarded to the Multiverz team, where these signals will undergo a final selection for the next level. The notification page will have all the requests(Refer Fig 29)

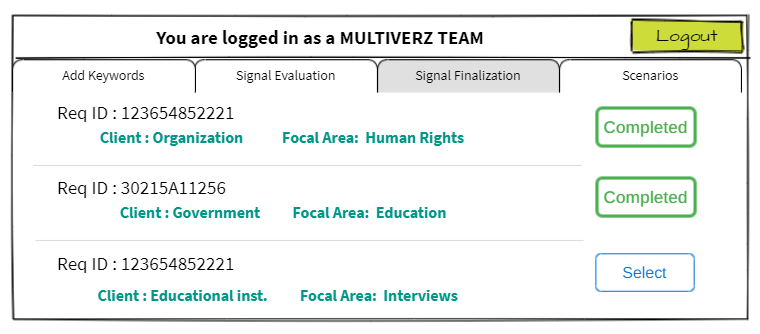


*Fig 29 : Notifications/requests for signal finalization(by Multiverz)*

Select the request which you want to work on, then all the details of the signals of the particular request will be shown (Refer Fig 30). Add the signals which have high impact and high complexity to the final signals and upon clicking Finish final validation, all the signals will be finalized. The status is updated in the notifications page as completed(Refer Fig 31)

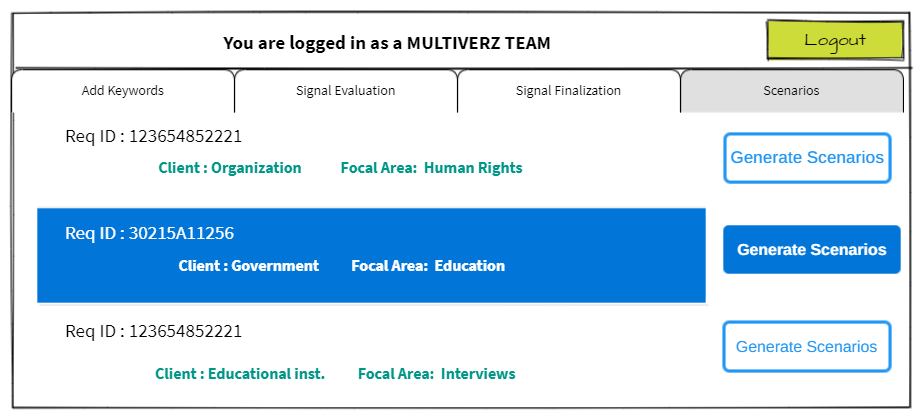


*Fig 30 : Signal Finalization(by Multiverz)*

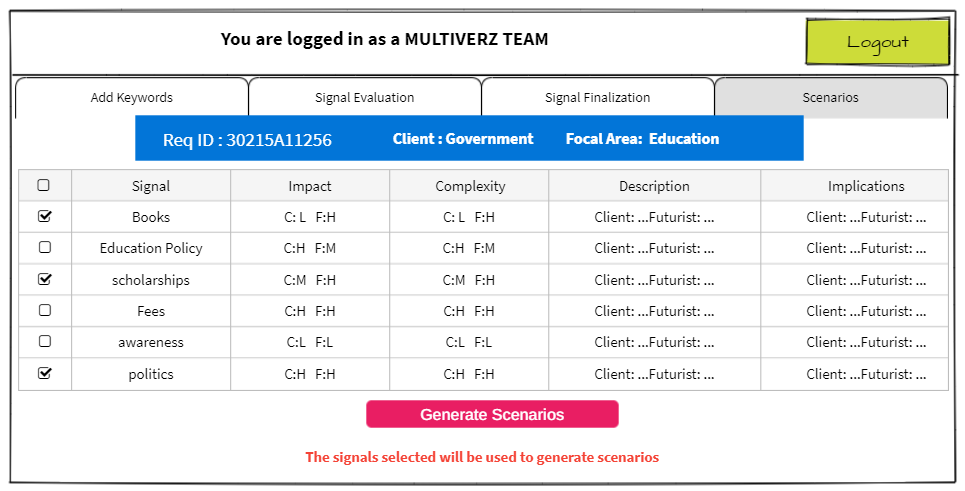


*Fig 31 : Signal evaluation completed status(by Multiverz)*

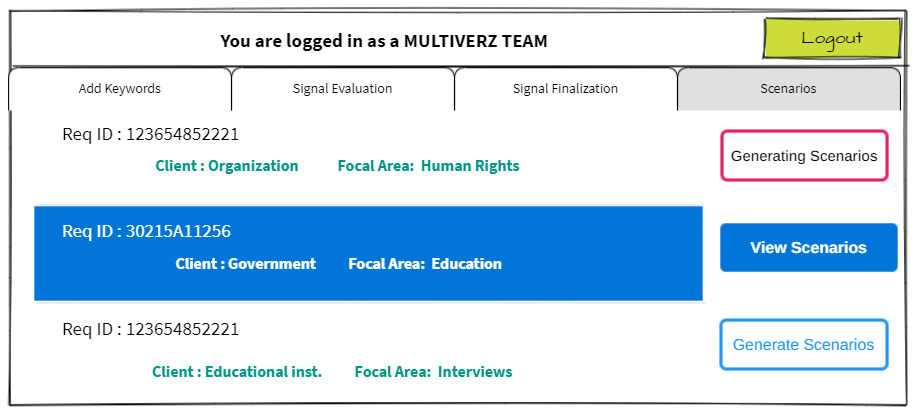
### Scenario Generation and Finalization

After the finalization of signals, Scenarios will come into picture. All the requests which completed finalization of signals stage will be updated in the Scenario notification page(Refer Fig 32. This Step is again done by the Multiverz team. Once the particular request is selected, a signal page will be displayed to the user with all the finalized signals and the team can select the signals which should be sent to GPT-3 as input to create scenarios(Refer Fig 33). Based on the signals selected in this stage, scenarios are created. 

*Fig 32 : Notifications/requests for Scenarios(by Multiverz)*



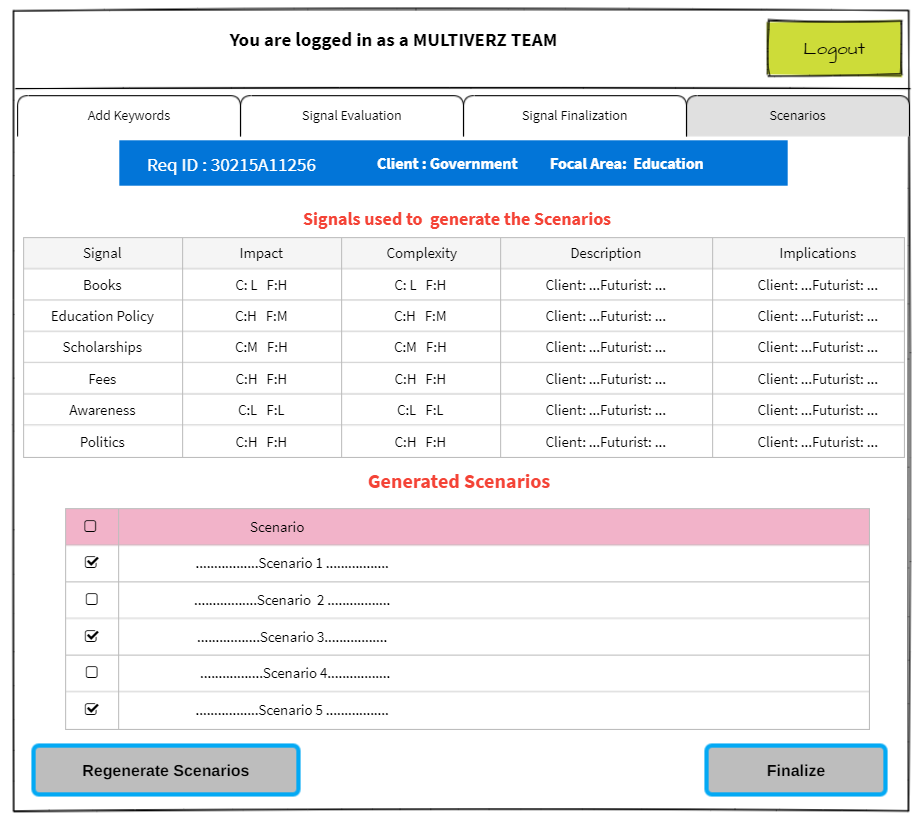
*Fig 33 : Signal selection for Scenarios(by Multiverz)*



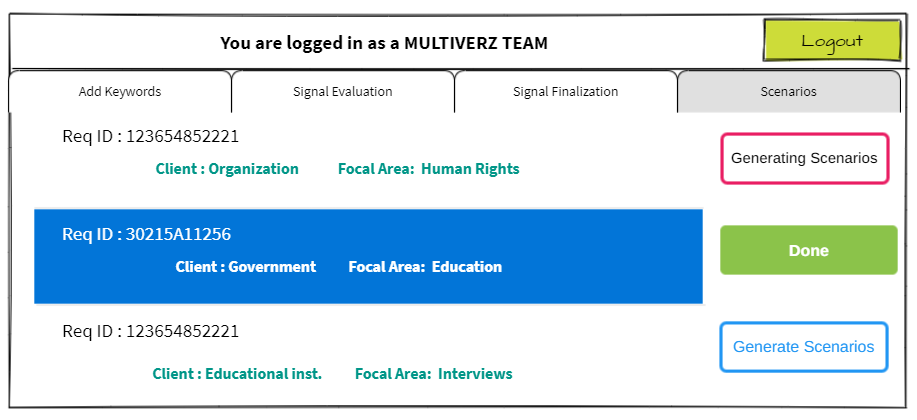
*Fig 34 : Notifications/requests for Scenarios(by Multiverz)*

In the backend the scenarios will be generated and the same information will be shown to the user(Refer Fig 34). Now, once the Scenarios are generated, the user can see the View Scenarios option and can click on it to see the Scenarios. All the generated signals along with the signals used to generate them are displayed to the user(Refer Fig 35). If the user is not satisfied with the results, there is a Recreate Scenarios option, which will take the user back to the signal selecting page(Refer Fig 33). If the user is satisfied with the scenarios generated and can select the scenarios which are to be finalized. Once the user clicks on Finalize, no further changes are encouraged and the request will be closed and the status of that request will be changed to Done(Refer Fig 36).

All the finalized scenarios are then sent to the admin, who first initiated and created the request. From the admin requests page, the admin can see the scenarios in the status of the request option.



*Fig 35: Scenarios generated for selected signals(by Multiverz)*



*Fig 36 : Notifications/requests for Scenarios(by Multiverz)*

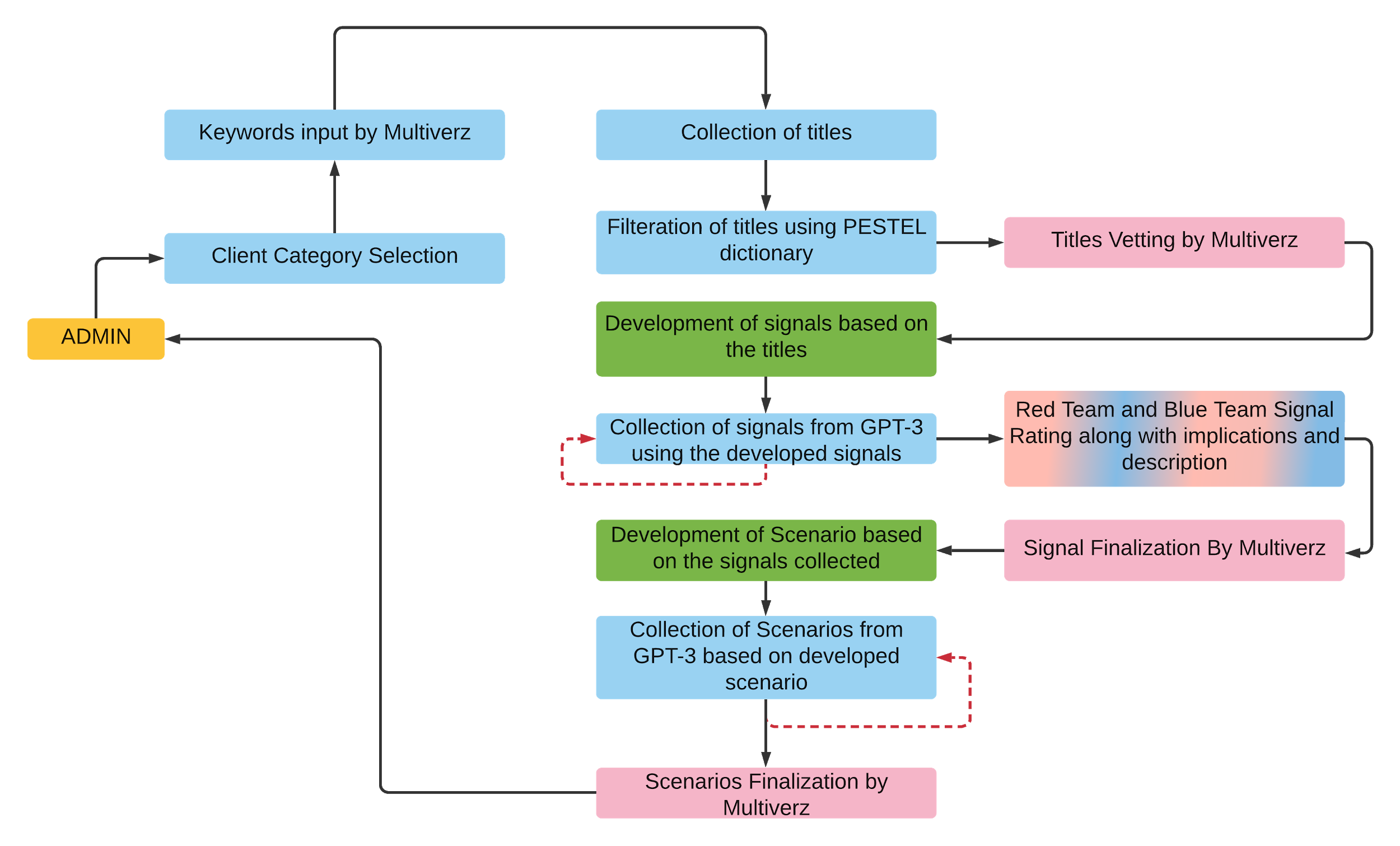
Approach 2

As in the first methodology, 4-5 resources(google search, patents, VC firms, social media, news) are used to extract the initial data required to start the process. This resulted in a huge size of the corpus, due to which it takes a lot of time to process when all the resources are included. Also it involves a lot of human effort in the further stages of evaluation and finalization.

Hence an alternative has been proposed which reduces the human effort by reducing the corpus size and improving the quality of signals that were received. A comparison of both the methodologies are shown in the table below.

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Approach 1** | **Approach 2** |
| Corpus size and predictability | * Large(517417) and Unpredictable | * Small(3300) and Fixed [3600] |
| Diversity in collected data | * High, as 4-5 resources are considered | * Only 1 resource is considered |
| Completion Status | * End to end implementation is not yet ready. | * End to end implementation is ready |
| PESTLE categorization count | P: 9120 Ec: 42116 S:11585 T: 21639 L: 55773 En:15263 | P:75 Ec:189 S:982 T:58 L: 172 En:214 |
| Time taken to extract the data | 5 ~ 5.5 hours(all resources considered)  3 ~ 3.5 hours(excluding google search with abstract)  30~ 60 min(excluding google search) | 30~ 60 min |

In this new approach the primary source of data is from the GDELT, which was used to collect the news articles in the first methodology. The working and the implementation of the GDELT was explained separately, refer that section for further details on the same.



*Fig 37:Complete process flow diagram of the second approach*

# Client Category Selection

As in the first methodology, the process starts with the raise of the request by the admin. The admin gives the area of interest(Focal area), in which the scenarios should be generated at a later stage. Along with the area of interest, the admin also mentions the client category. The outputs from this stage i.e., the focal area and client category will be fed as input to the next stage.

# Keyword inputs from Multiverz

Based on the inputs received from the previous stage, the Multiverz team will provide inputs in the form of keywords which will then be used to initiate the model.

# Collection of Titles

The input keywords given in the previous step will be considered as the query, which is used to start extracting all the related articles from the GDELT with the help of API. With each API request a maximum of 75 articles can be extracted from time period 2017 to till date. In the implementation, the 12 requests are made to collect data monthly on a yearly basis from 2017. The time range has been limited to 2020 in the implementation, which can be modified as per the requirement. Hence by the end, a total of 3600 (75\*12\*4) articles are extracted in the ideal case i.e., each request will return 75 articles depending on the availability of the articles in the GDELT database. This again depends on the input query.From the articles all the titles are collected. As the response is in the Json format(other formats are also available), it will be easier to collect the title directly from the response object.

For the query “Fashion Future” 3300 articles were collected. It will take approximately 20-30 minutes to complete the extraction process. Time taken again depends upon the availability of the articles in the GDELT database.

# Filtration of Titles using PESTEL dictionaries

As the number of titles collected are huge in number and in order to filter the titles that belong to either of the categories in PESTEL, this stage is required. The titles which are collected from the GDELT are filtered using the PESTLE word banks. These word banks are the curation of the 6 dictionaries which represent the categories of PESTEL.

In this step, the category of each title is identified and is tagged under that particular category. One possibility in this case is that a particular title may come under different categories. In such cases the respective title will be tagged under multiple categories.

# Titles Vetting by Multiverz

Upon categorization of the titles, the first stage of vetting happens by the Multiverz team, who select the titles which can be used in the further steps from all the six categories of PESTLE. At this stage, human intervention is required to pick up the important or significant titles.

# Development of signals based on Titles

The titles that were selected by the Multiverz team in the previous step are shown to a futurist. Based on the titles, the futurist will develop a couple of signals. This is the important step in the entire process, because these signals will be fed as input to the GPT-3 to generate more and more signals. The signals are to be curated carefully, as the rest of the signals which will be generated will be of the same template or format. A minimum of 2 signals will be efficient to get reasonably accurate signals. Anything more than that will also be useful in getting the signals, as expected.

# Collection of signals from GPT-3

The signals curated by the Futurist will act as input to the GPT-3. GPT-3 working and implementation are explained in earlier sections separately. The new signals from GPT-3 i.e., the output is collected in an iterative manner, which means the output received from the GPT-3 is collected and combined with the original input and that input is again fed to GPT-3 to generate new results. This process continues for 3-4 times until a reasonable number of signals are received at the end. The output from this stage are the signals related to the focal area, as per the request raised by the admin.

# Evaluation of Signals by Red Team and Blue Team

This is the second evaluation that involves human intervention in the entire process. The signals generated by the GPT-3 are evaluated by two teams separately. The Red team is the client team and the Blue team is the Futurists. Both the teams will be provided with the same signals and each signal should be rated on two different parameters Impact,Complexity on a scale of 3(low, medium, high). Along with the ratings, the description and implications of the signal will also be written by each of the team. Once both the teams finish their evaluation, the signals will be taken ahead for the next steps.

# Signal Finalization

By considering the ratings given by the two teams, Multiverz team will select the signals. The signals selected by the team will be finalized, which are then considered for the scenario generation.

# Development of Scenarios

Based on the signals finalized, futurists will write some scenarios. These scenarios are to be fed to the GPT-3. The end output in this entire process is dependent on the scenarios written.

# Collection of Scenarios from GPT-3

From the GPT-3 scenarios are generated based on the scenarios written by the futurists in the previous stage. The method used to generate the scenarios will be similar to that of the method used to generate the signals from GPT-3 in the previous stages. The scenarios will be generated in an iterative manner with a minimum of 3 iterations, which can be increased further depending on the output quality.

# Scenario Finalization

All the scenarios that are generated by the GPT-3 are displayed to the Multiverz team. The team will select the scenarios and all the selected scenarios will be finalized and which are then forwarded to the Admin, who initiates the request. After this, the initiated request will be closed.

References

1. GDELT: [GDELT Summary: Online News (gdeltproject.org)](https://api.gdeltproject.org/api/v2/summary/summary)

### 