

## **Team 50**

Members:

- Shashwat Singh
- Venika Shruti
- Tisha Dubey
- Lokesh Paidi

## **Design Overview**

### **Architectural design**

#### **Tf-idf vectorization model**

It has been assigned with the part that it will put the news articles into one of the two bins:

- Related to import/export :
  - If it is related to import/ export it will be sent to the database.
- Not related to import/export:
  - It will be removed from consideration.

The Tf-idf model has been trained on the corpus of Maritime gateway and BBC datastore(to train the the negative corpus).

### **Keyword extraction**

It will reduce each article into a queryable set of words i.e., it will rank the article based on the some keywords such as country specific where it is exporting and importing.

## **System Interfaces**

### **APIs**

User queries a feed and sends their token to receive user specific relevant data.

## Model

Class State	Class Behavior
Article will be scraped; has title, author name and relevant information , vectorized form	Methods: vectorize, string search
User: data about business, export area and products	getters and setters

## Sequence Diagram

### Design Rationale

- The article has been scraped to get titles, author name and relevant information.
- The article has been taken from the Maritime gateway and BBC
- If the article is related to import/export it will be stored in a database.
- The API will fetch the query from the database.
- User data will be available from the API provided by the client.
- For the ranking we'll need to do string matching on the keywords.