Analysis of Global Air Route Networks Using Machine Learning

Arhit Bose Tagore
Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
arhitbose.tagore2020@vitstudent.ac.in

Candace Joy Pals
Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
candacejoy.pals2020@vitstudent.ac.in

Megha Jayantilal Bhimani
Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
meghajayantilal.bhim2020@vitstudent.ac.in

***Abstract*—This study explores the pivotal role of Indian airports in international relations by analyzing the quantity of flights connecting them to various states. Our focus is to identify the Indian airport serving as a key hub, facilitating extensive international connectivity. Additionally, a global airline itinerary map is developed to visually represent the intricate network of flight paths worldwide. To bolster our analysis, a sentiment analysis model trained on a US dataset is employed, providing insights into the diplomatic and economic implications of international flights. By combining quantitative assessments of flight volumes with sentiment analysis, we aim to unveil the Indian airport with the strongest international relationships, contributing to a nuanced understanding of the role airports play in fostering global connections and the dynamics shaping international air travel.**

***Index Terms*—Air Network Topology, Homophily in Air Networks, Covid-19 Impact on Aviation, Degree Centrality, Antisocial Nations, Global Commerce Routes**

I. Introduction

The symbiotic relationship between a nation's economic prosperity and the vitality of its transportation and trade sectors is an established paradigm. At the forefront of this nexus are airports, serving as gateways to international connectivity. This paper delves into the intricate dynamics of global air networks, employing novel approaches to gain insights into the economic, social, and geopolitical dimensions intertwined with air travel.

Our exploration begins by mapping the nodes of these networks onto a world map, offering a visual representation that allows for a nuanced understanding of global connectivity patterns. Recognizing social media as symbols of homophily, we conduct a case study focusing on countries like North Korea and China, drawing analogies between their airline networks and national philosophies. This comparative analysis sheds light on international commerce routes and diplomatic ties.

Dispelling the notion that "more is better," we scrutinize the dense air network of the United States, investigating traveler satisfaction to discern the quality-quantity correlation. Moreover, our inquiry extends to the repercussions of the Covid-19 pandemic on the aviation industry, emphasizing the profound impact of global events on air travel.

Key questions guiding our investigation include identifying the nation with the most airports, assessing the correlation between a nation's airport count and its developmental level, scrutinizing air routes to antisocial nations, and unraveling the nuanced relationship between quantity (degree centrality) and quality in air networks. This multifaceted exploration aims to contribute to a comprehensive understanding of the intricate interplay between air travel, socioeconomic factors, and global events.

II. Methodology

Our research employs a multi-faceted methodology, combining data visualization, social network analysis, sentiment analysis, deep learning models, and comparative mapping to address the key problem statements and unravel the complexities of global air networks.

A. Data Preprocessing and Visualization

We preprocess acquired data sets to represent every airport node on a global map. Utilizing globe-based visualization, we establish connections between airports with direct flights, providing a comprehensive image of the global airline network.

B. Social Network Analysis

The study delves into the concentration of airports across countries, ranking them by quantity. A comparative case study between North Korea and the United States offers insights into the variations in aviation network density.

C. Sentiment Analysis through Tweets

Conducting sentiment analysis on tweets enables a quantitative and qualitative assessment. We investigate whether a nation with a high airport count corresponds to a robust airline presence, with a focus on the United States.

D. Machine Learning for Satisfaction Analysis

Employing a deep learning model, we analyze sentiment in the United States' tweet dataset to gauge public satisfaction with domestic air travel, probing beyond mere quantitative indicators.

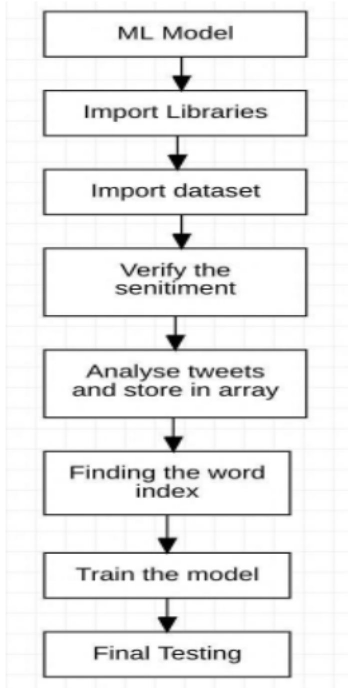

Fig. 1. Flow diagram of the ML model

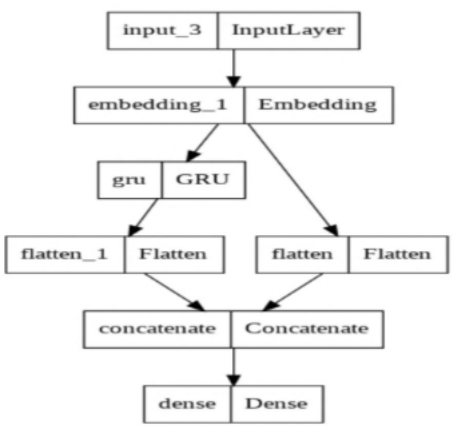

Fig. 2. ML Model of the dataset analyzing

E. Proximity Centralities and COVID-19 Impact

Mapping proximity centralities for each nation, we correlate COVID-19 instances during the early pandemic stages with the global air network. A comparative graph illustrates the pandemic's impact, detailing scheduled and cancelled flights.

This comprehensive methodology aims to provide a holistic understanding of the intricate relationships between airport quantity, network density, public satisfaction, and the global air network's resilience to external shocks like the COVID-19 pandemic.

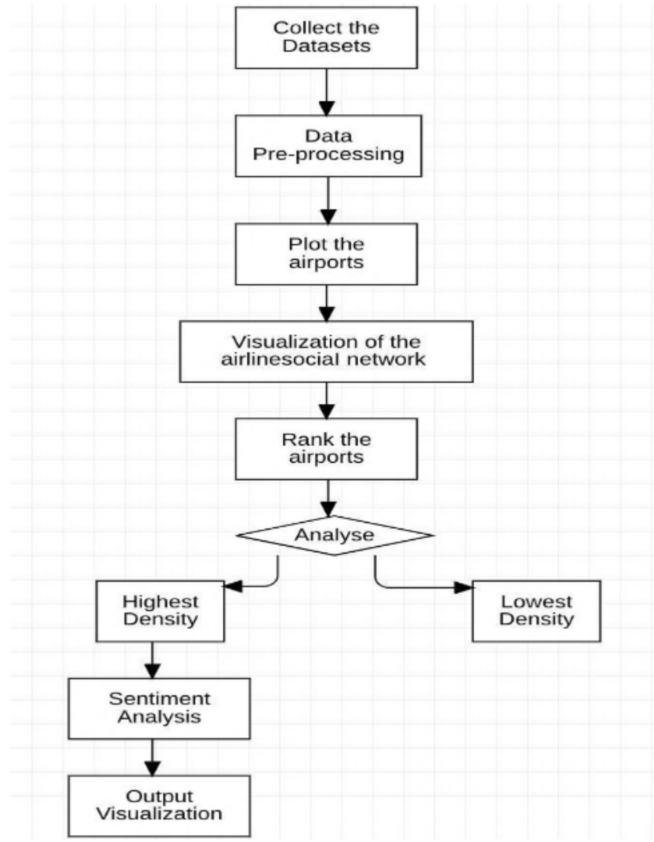

Fig. 3. Flow diagram of the project

III. Module Description

A. Module 1: Plot-Air Routes

Description: The Plot-Air Routes module plays a pivotal role in the project's initial stages, focusing on the visualization and representation of global air routes as a social network. This module involves the processing of acquired datasets, specifically those detailing airport locations and flight routes. Through careful preprocessing, the data is structured to highlight direct flight connections between airports. The module then utilizes globe-based visualization techniques to plot these airports on a world map, creating an intricate network of connections. This visual representation not only aids in understanding the global air network's complexity but also serves as a foundation for subsequent analyses.

Key Functions:

- **Data Processing:** Clean and structure datasets related to airport locations and flight routes.
- **Network Plotting:** Utilize geographic information to create a visual representation of airports and their connections on a world map.
- **Graphical Visualization:** Employ network visualization techniques for enhanced comprehension of global air routes.
- **Iterative Data Exploration:** Analyze connected nodes iteratively to capture the nuances of the global airline

network.

Output: A visually intuitive representation of global air routes, providing a foundation for further investigations into the socioeconomic and geopolitical aspects of the aviation industry.

B. Module 2: Sentiment Analysis

Description: The Sentiment Analysis module is designed to gauge the sentiments of customers engaged in air travel through the analysis of tweets. Central to this module is a machine learning model trained on a dataset of tweets related to air travel. The model systematically assesses the sentiment expressed in these tweets, providing valuable insights into public perceptions and opinions regarding airport services and air travel experiences.

Key Functions:

- **Tweet Collection:** Gather a diverse dataset of tweets related to air travel, considering factors such as origin, destination, and airline.
- **Machine Learning Model:** Train a sentiment analysis model using machine learning techniques to classify tweets into positive, negative, or neutral sentiments.
- **Quantitative vs. Qualitative Analysis:** Explore whether nations with a higher number of airports also exhibit a proportionate level of positive sentiment in tweets related to air travel.

Output: A quantitative and qualitative analysis of public sentiment towards air travel, offering valuable insights into customer satisfaction and the correlation between sentiment and airport quantity.

IV. Software Requirements for the Project

1) **Python (Version 3.x):**
   - **Description:** Python is the primary programming language for data processing, analysis, and implementation of machine learning models.
   - **Justification:** Many data science and machine learning libraries are readily available in Python, making it a preferred choice for data manipulation and analysis tasks.
2) **Jupyter Notebooks:**
   - **Description:** Jupyter Notebooks provide an interactive computing environment for code execution, data visualization, and documentation.
   - **Justification:** Jupyter Notebooks facilitate a collaborative and iterative development process, allowing for a seamless integration of code and visualizations.
3) **Pandas:**
   - **Description:** Pandas is a powerful data manipulation library in Python, essential for cleaning, processing, and organizing datasets.
   - **Justification:** The Plot-Air Routes module relies on efficient data processing, making Pandas an integral component for handling tabular data.
4) **Matplotlib and Seaborn:**
   - **Description:** Matplotlib and Seaborn are Python libraries for creating static, interactive, and aesthetically pleasing visualizations.
   - **Justification:** These libraries are essential for plotting global air routes and generating insightful visual representations.
5) **Natural Language Toolkit (NLTK):**
   - **Description:** NLTK is a comprehensive library for natural language processing and sentiment analysis.
   - **Justification:** NLTK is employed for sentiment analysis in the Sentiment Analysis module, providing tools for text processing and sentiment classification.
6) **Scikit-Learn:**
   - **Description:** Scikit-Learn is a machine learning library in Python that provides simple and efficient tools for data analysis and modeling.
   - **Justification:** Scikit-Learn is used to train machine learning models for sentiment analysis on the Tweet dataset.
7) **Deep Learning Framework (e.g., TensorFlow or PyTorch):**
   - **Description:** A deep learning framework is required for training and deploying deep learning models for sentiment analysis.
   - **Justification:** Deep learning models enhance the accuracy of sentiment analysis by capturing complex patterns in textual data.

V. Datasets Used

A. Statistics of Passengers for All Carriers

- **Description:** This dataset provides comprehensive statistics on passengers for all carriers, offering valuable insights into air travel patterns, carrier-specific performance, and overall industry trends. The dataset includes information on the number of passengers, flights, and other relevant metrics.

B. Airports and Airlines Data

1) OpenFlights Data:

- **Description:** The OpenFlights dataset is a rich source of data on global airports and airlines. It includes geographic information, routes, and additional details that contribute to a comprehensive analysis of the aviation landscape. This dataset is crucial for mapping and understanding the global network of airports and airlines.

2) OurAirports Data:

- **Description:** The OurAirports dataset complements the information provided by OpenFlights, offering additional details on airports globally. This includes information on airport types, runways, and other relevant attributes, contributing to a more nuanced understanding of the aviation infrastructure.

C. Twitter US Airline Sentiment

- **Description:** The Twitter US Airline Sentiment dataset, available on Kaggle, comprises tweets related to US airlines. This dataset is instrumental for sentiment analysis, allowing for an exploration of public opinions and attitudes towards air travel services. It includes sentiment labels and other metadata associated with each tweet.

VI. Equations Used

A. Equation 1: Network Centrality Calculation

The centrality of a node in the air network is calculated using the degree centrality formula:

$$C_D(v) = \frac{\text{Number of connections of node } v}{\text{Total number of nodes in the network}} \tag{1}$$

Where:

$$C_D(v) : \text{Degree centrality of node } v$$

B. Equation 2: Sentiment Analysis Model

The sentiment analysis model employs a deep learning approach to classify sentiment in tweets. Let X be the input features and Y be the predicted sentiment label. The model is trained using a neural network with parameters W and biases b:

$$Y = \sigma(W \cdot X + b) \tag{2}$$

Where:

$$\sigma : \text{Activation function (e.g., sigmoid or softmax)}$$

VII. Results

A. Global Air Network Visualization

The visualization of the global air network revealed insightful patterns and connections between airports. Figure **??** displays a comprehensive map of airport nodes and their interconnections.

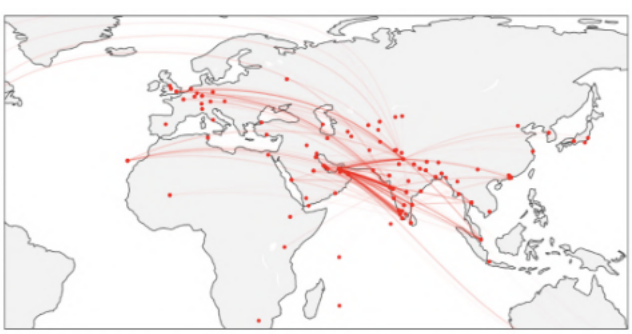

Fig. 4. Global Air Network Visualization

Fig. 5. All Airlines Visualization

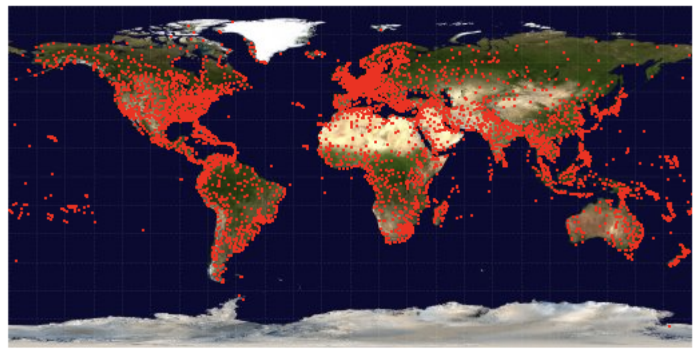

Fig. 6. All Airports Visualization

B. Degree Centrality Analysis

The analysis of degree centrality provided valuable insights into the prominence of airports in the network. Table I presents the degree centrality values for selected airports.

Airport	Degree Centrality
FRA	0.042946
CDG	0.042316
AMS	0.041685
IST	0.041145
ATL	0.038984
PEK	0.037094
ORD	0.036824
MUC	0.034213
DME	0.034033
DFW	0.033492

TABLE I
DEGREE CENTRALITY ANALYSIS

C. Sentiment Analysis Results

Sentiment analysis of tweets related to air travel yielded interesting findings. Figure 7 illustrates the distribution of sentiments across different airlines. Our research employs a multi-faceted methodology, combining data visualization, social network analysis, sentiment analysis, deep learning models, and comparative mapping to address the key problem statements and unravel the complexities of global air networks. Refer Fig. 7

Airlines present in US and their share market

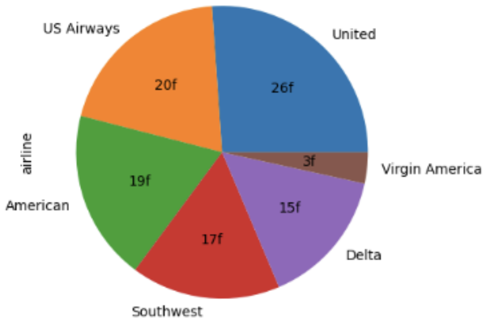

Fig. 7. Airlines present in US and their share market

Comparison between share of each sentiment: positive, negative and neutral. Overall sentiments in our United States flight dtset (independent of irIine) – Post sentiment analysis we see that despite the biggest and the most robust flight network, the overall sentiment of the fliers is still negative.

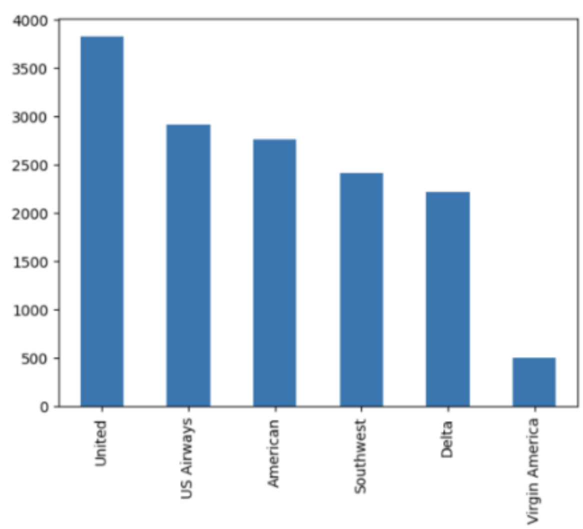

Fig. 8. Airlines present in US and their share market

Airlines present in US and their share market

irIine vs the kinds of sentiment ssocited with them.- Some of the popular US airlines and the general sentiment the fliers have towards them

VIII. Future Work

Following the outbreak, demand for flights has increased across the globe for airlines. Thus, in order to make educated judgements and forecasts, an increasing number of airlines are searching for effective data analysis. To improve efficiency and optimise fuel costs for airlines, the project might be expanded to include vehicle routing issues for connecting airlines. Additionally, by examining the most popular tweets on Twitter, SEOs can be achieved and pre-planned marketing campaigns can be implemented. The project can be expanded further to create flight schedules for the employees based on the routes they take during flying hours and shift times. Colleagues' work environment can be developed with the aid of a thorough investigation of the work

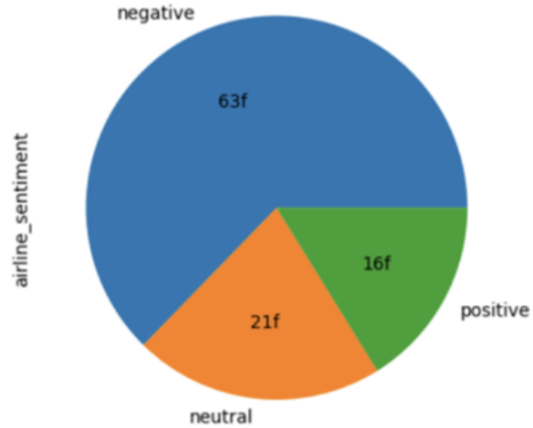

Fig. 9. Airlines present in US and their share market

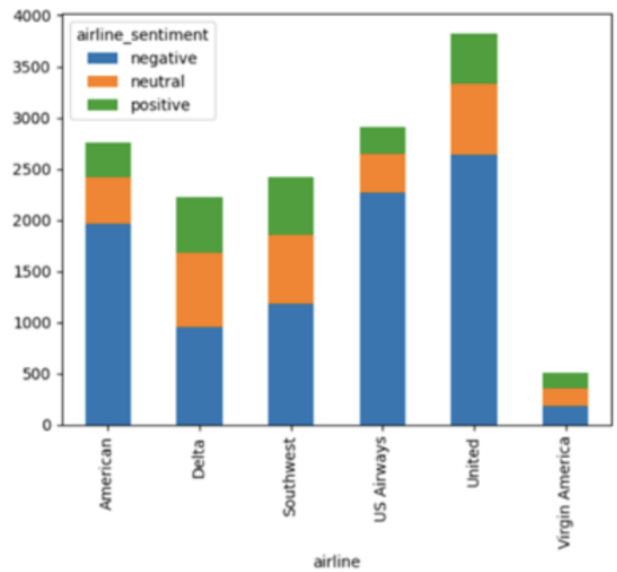

Fig. 10. Airlines present in US and their share market

IX. Conclusions

A comprehensive overview of the global airline network has been provided. The world's airports will be represented as nodes on a map of the globe, with the airport's latitude and longitude serving as the coordinates. An equivalency of the number of airports in each nation is achieved through a visual comparison. A graph showing the number of direct flights that operate between various nations.

There will be a sentiment analysis done, classifying the tweets regarding US airlines as neutral, negative, or favourable. Given that the US has the greatest concentration of airline services globally, it is reasonable to assume that the majority of Americans are satisfied with their flight experiences. This anticipation has been proved correct. Sentiment analysis demonstrates that quality is not always correlated with quantity, since the majority of US passengers are dissatisfied with the nation's air travel offerings. International flight travel was one of the main causes of the sudden increase in Covid 19 cases in several countries.In the past year, the air travel industry has experienced an exponential fall due to the epidemic

REFERENCES

[1] Aura Reggiani, Peter Nijkamp, Alessandro Cento. (2010). "Connectivity and concentration in airline networks: a complexity analysis of Lufthansa's network." *European Journal of Information Systems,* 19(4), 449-461.

[2] Zanin, Massimiliano, and Fabrizio Lillo. (2013). "Modelling the air transport with complex networks: A short review." *The European Physical Journal Special Topics,* 215, 5-21.

[3] Bounova, Gergana, et al. (2006). "Analysis and Optimization of airline networks: A Case Study of China." *Santa Fe Institute of Theoretical Physics and Chinese Academy of Social Sciences CASS.*

[4] Suzumura, Toyotaro, et al. (2020). "The impact of COVID-19 on flight networks." *2020 IEEE International Conference on Big Data (Big Data).*

[5] Calderon-Tellez, Javier Andres, and Milton M. Herrera. (2021). "Appraising the impact of air transport on the environment: Lessons from the COVID-19 pandemic." *Transportation Research Interdisciplinary Perspectives,* 10, 100351.

[6] Adiga, Aniruddha, et al. (2020). "Evaluating the impact of international airline suspensions on COVID-19 direct importation risk." *medRxiv.*

[7] Serrano, Francisco, and Antonín Kazda. (2020). "The future of airports post COVID-19." *Journal of Air Transport Management,* 89, 101900.

[8] Agrawal, Anshu. (2021). "Sustainability of airlines in India with Covid-19: Challenges ahead and possible way-outs." *Journal of Revenue and Pricing Management,* 20(4), 457-472.

[9] P, S., Swetha, P. K., & Radha, D. (2018). "Analysis and visualization of airport network to strengthen the economy." *International Journal of Engineering & Technology,* 7(2), 708–713.

[10] Sun, Xiaoqian, Sebastian Wandelt, and Anming Zhang. (2020). "How did COVID-19 impact air transportation? A first peek through the lens of complex networks." *Journal of Air Transport Management,* 89, 101928.