UDC 004.94

D. Tarasonok, T. Likhouzova

MODELS FOR RESEARCHING PROSPECTS FOR THE DEVELOPMENT OF THE AVIATION INDUSTRY

Abstract: Considered the problem of prospects for the development of the aviation industry in the world. It is proposed to use machine learning models and technologies for clustering countries by the number of air flights and forecasting the production of aircraft in the future. Several different quality measures were used to evaluate the effectiveness of the proposed models, which reflect the needs of each task in various ways.

Keywords: machine learning, prediction of time sequences, regression problem, clustering problem.

Introduction

Currently, humanity cannot imagine its existence without air travel. Such a large industry simply needs to analyze trends in detail to improve logistics and increase profits.

So, for example, the clustering of countries will allow airlines to study in detail the possible directions of expanding their business, to look at the main centers of aviation, to see which countries they can expand their business to.

A detailed analysis of trends in aircraft production is the main thing that aircraft manufacturers should focus on since such trends change very quickly and one must try to predict them. Not long ago, the world celebrated the appearance of such a giant as the Airbus A380, and already in 2019, the Airbus company announced that it would stop production of aircraft of this model by 2021 [1]. This is due to a number of factors: the aviation industry has begun to move away from the concept of a hub and to operate shorter, non-stop flights on a larger scale. Also, the world is starting to get rid of 4-engine aircraft, because they were once necessary for greater reliability, and now the 2-engine Boeing 777 is considered the best and most reliable engine in the history of aviation.

All these trends are very important to monitor and predict for both airlines and aircraft manufacturers, so as not to repeat the fate of the Airbus A380 - a large, beautiful, powerful, but useless aircraft.

Within the framework of this work, the aviation industry was analyzed to obtain a comprehensive picture of the present and the future of this field. Two tasks were set: clustering of countries in order to identify new promising routes and forecasting the production of various aircraft. For the first task, a clustering model was built, a graph was created from it, and this graph was displayed in a web format for easy interpretation of the results. For the second task, several forecasting models were created for different categories of aircraft and the forecast of aircraft production for the coming years was performed.

Python 3 packages were used to perform the tasks: scikit-learn, pandas, NumPy, Matplotlib, Keras [2], python-louvain [3], Prophet [4], the Gephi software application [5] and the nginx web server [6].

Materials and methods

Using archival and secondary data, mainly from Flightradar24, ICAO, IATA, and EUROCONTROL, the study [7] found that the pandemic took a heavy toll on global aviation, leading to the downgrading, liquidation, and bankruptcy of several airlines and airports due to severe monetary losses caused by travel restrictions. The study tracked the impact and recovery paths of the global aviation industry following the COVID-19 pandemic and primarily focused on North America, Europe, and Asia Pacific.

Prolonged travel restrictions around the world have put the industry in a precarious position, with several airlines downgraded, raising restart and capitalization costs. The pandemic offered aviation the opportunity to start a new life, based on operational efficiency and technological progress.

Over the past few decades, the aviation industry has created several innovative solutions that have benefited the airline industry and many other industrial and service organizations. However, few attempts have been made to outline these innovations and, more importantly, to assess how they add value to the aviation business. The article [8] provides a systematic literature review (analyzed 57 peer-reviewed academic publications from 1999 to 2018) to analyze growth points, approaches and how innovation creates value in this sector. The results identified ten key business growth areas for innovation: airports, services, management, technology, airlines, systems, development, market, customers and processes.

In [9], 130 articles published between 1993 and 2020 are reviewed to summarize the literature dealing with the specific application of DEA (data envelopment analysis) models to evaluate airline performance.

Data loading and preprocessing

The process of collecting, preparing, and downloading data for the given task is described in detail in [10]. This data includes information about altitude, coordinates, departure and destination airports, airlines, aircraft models, and flight numbers - that is, complete information about every aircraft in the sky, as well as the meteorological conditions under which the flight was carried out. Information on aircraft models from the ICAO website [11] and on airports, country regions, and countries [12] was also downloaded. Of these, 562,894 flight numbers and 278,508 aircraft registration numbers were exported.

Data preparation for clustering

For clustering, the data set was grouped by departure airport and destination airport and the number of such flights was counted. The source file contains more than 490 thousand

lines. This data set had to be supplemented: with information about airports (only the region code), about regions (only the country code), and about countries (only the name of the country), after which the region and country code were deleted, the data were grouped by countries instead of airports, the number of flights was calculated for each country and removed information about airports. We can depict the obtained data set on graphs (Fig. 1, Fig. 2).
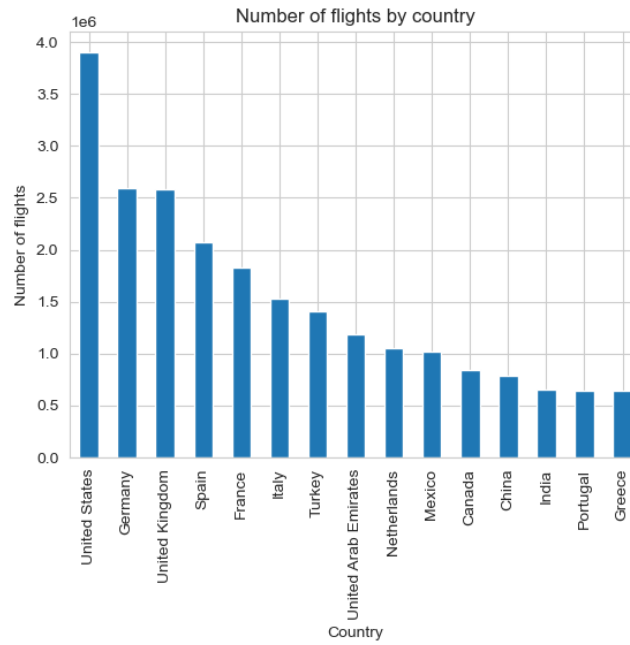

Fig. 1. Total number of flights by country

Preparation of data for aircraft production forecasting

Aircraft information has been supplemented with aircraft model information from the ICAO website. The merging process was as follows: the aircraft data set, as already mentioned, contained information about the model code and the model text. The model data set was sorted by model code, all model texts were taken (because different aircraft models can be under the same code) and then calculated using the sequence comparison algorithm presented in the difflib library of the Python language as a modification of the algorithm by Ratcliffe and Obershelp under called "matching Gestalt patterns" to compare the texts of the models, after which the best match was taken.

Information about aircraft models is described by the following parameters [10]: manufacturer, model name, model code, aircraft type, engine type, number of engines and its turbulence category. Under one model code, there can be a large number of different modifications of this model. Currently, aircraft are manufactured with the following types of engines: electric, jet, piston, rocket, and turboprop (or turbocharger). The turbulence category is a category that actually indicates the weight and size parameters of the aircraft. Can have

the following values: L (small), L/M (small-medium), M (medium), H (heavy), and J (Airbus A380 only). In fig. 3 visualizes the share of aircraft models in the overall structure of flights.
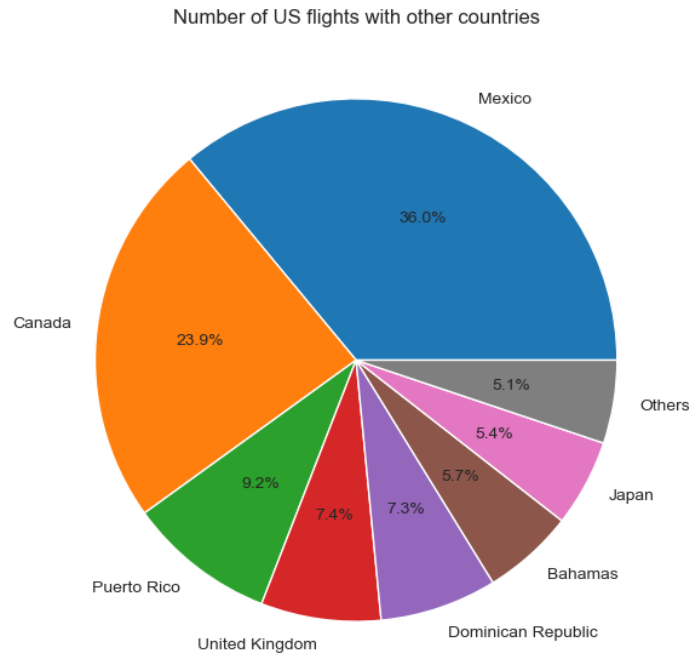
Number of US flights with other countries

Fig. 2. Shares of countries in air traffic with the USA
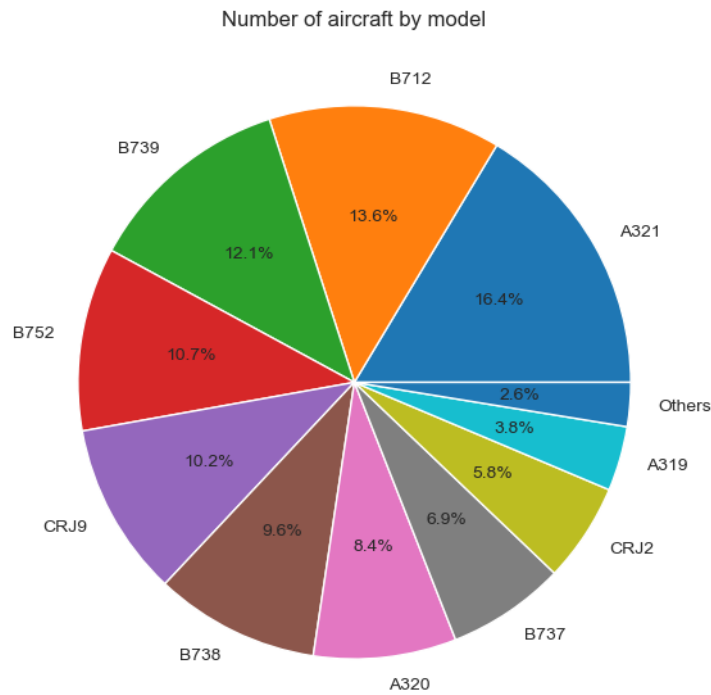
Number of aircraft by model

Fig. 3. Share of aircraft models in the overall structure of flights

The schedule of total aircraft production by year is shown in Fig. 4.
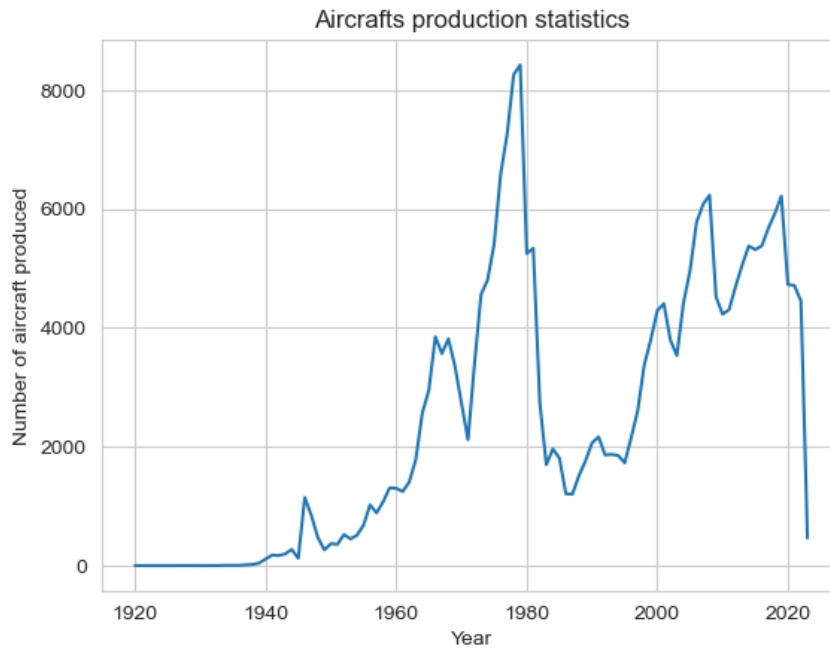

Fig. 4. Schedule of total aircraft production by year

As we can see on the graph, in 2020 there is a sharp decline in aircraft production. The reason for this is very simple: the coronavirus pandemic, which caused a significant drop in production. As of 2023, the indicators are returning to normal, but since our goal is long-term forecasting, data for 2020-2023 has been discarded.

Results and discussion

The task of country clustering

The Louvain method was chosen to solve the task of clustering countries. When choosing a model, we were inspired by the study of clusters of countries in arms circulation. Currently, this method is the fastest among all analogs, according to the article by its developers themselves [13]. Compared to some algorithms, it works hundreds or even thousands of times faster.

When preparing the data, a set was formed that contains the same names of columns and rows - countries and at the intersection has the number of flights between these countries. First, we will make a graph in Python, and then we will simply perform clustering using the python-louvain package. Let's add the cluster number to the data set and display a graph showing the number of countries belonging to each of the clusters (Fig. 5). As you can see, the distribution is quite uneven, but this does not indicate incorrect clustering.

A very nice visualization of this method is the image in the form of a graph with the colors of the clusters. Let's add some additional information to our graph: the total number of

flights for the country and its rank in the overall ranking, as well as the importance of each of the connections to the countries they connect and the connection's rank in the overall ranking.

Next, we can export this graph to a file and proceed to work with the Gephi program, in which we can configure the visualization of the graph. To place the nodes on the graph, we will use the Fruchterman-Reingold algorithm [14], having previously configured it for a beautiful display of the nodes on the graph. After creation, visualizations are exported in web format for viewing in a browser (they are located on the website at the address https://flights.compich.com/clustering/ [15] where they can be viewed in more detail).

The first visualization (Fig. 6) shows all the connections, which makes it quite difficult to see any information. The only thing you can get is to which cluster country belongs and to see the connections of the country by clicking on it. We also see a completely natural, but interesting thing: the USA is in the very center of the graph, as it has strong ties with most countries of the world.
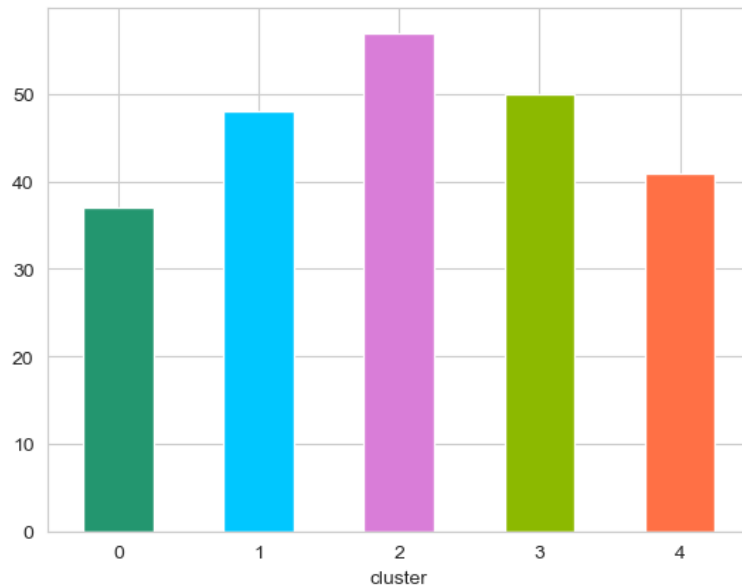

Fig. 5. Distribution of countries by clusters

The second visualization (Fig. 7) displays only the most important connections for each of the countries. Here you can see more detailed trends in movement between countries in the world, note which countries are more important for a particular state, as well as how people from one country get to another.

You can also see a very unusual thing here: Ukraine has the closest ties with the following countries: Turkey, Egypt, and Russia, but this is impossible even theoretically, since Ukraine has had no air connection with Russia since 2014, and the number of air flights to the aggressor country was measured in units. and information on FlightRadar24, as already mentioned, is stored exclusively for the last 3 years. The answer to the question actually lies

on the surface. The fact is that the international community considers Crimea to be Ukraine, and therefore exactly such information is contained in the data set of regions. This detail was not intentionally removed from the data set, since this clustering model was created to identify such nuances.

The following visualization (Fig. 8) depicts the 100 most important world connections. Here you can see that these are mainly either resort countries (mainly those to which Americans fly) or rich countries. Also, the main transfer hubs, where the world's largest airlines are based, are immediately highlighted: USA – American Airlines, Delta Air Lines, United Airlines; Germany – Lufthansa; Great Britain - British Airways, China - China Airlines, Air China; France – Air France; United Arab Emirates - Emirates.
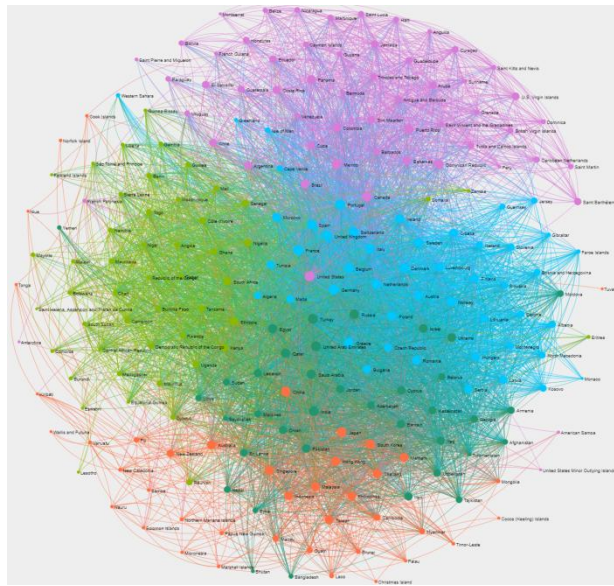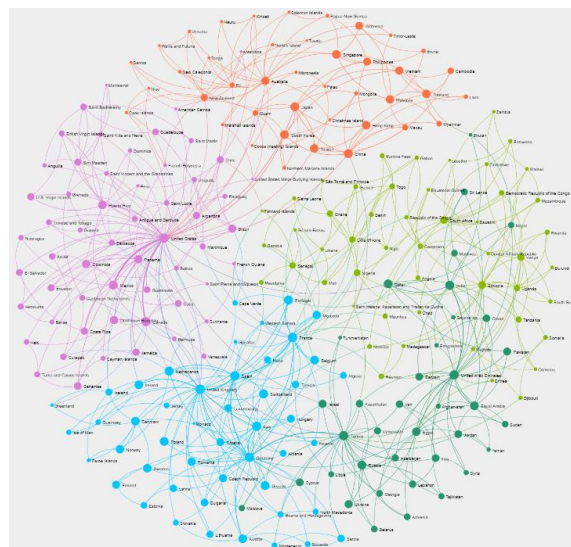

Fig. 6. Visualization with all connections

Fig. 7. Visualization with the three most important connections for each country
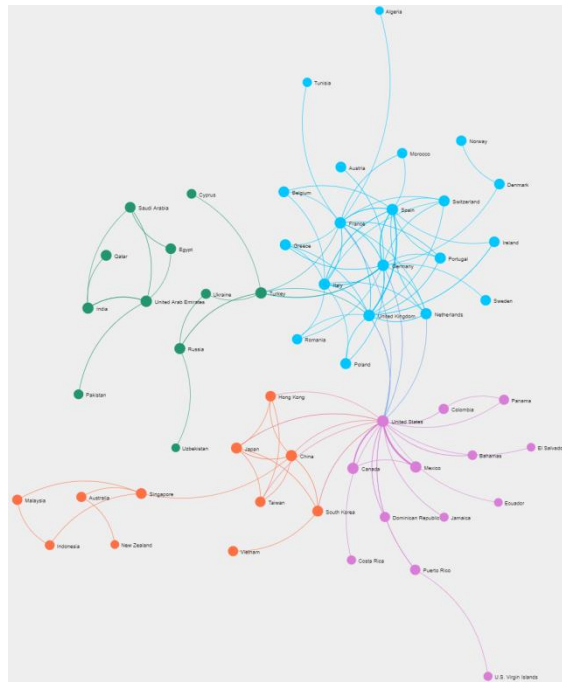
Fig. 8. Visualization with the hundred most important world connections

We can see that the cluster of Eastern European and Middle Eastern countries (dark green in the visualization) has broken. This was due to the fact that the filtering of connections took place already after clustering for greater accuracy and indicates the theoretical possibility of dividing this cluster into two smaller ones.

It is also important to remember that due to the lack of data on passenger traffic, it is necessary to analyze the number of air flights, which cannot accurately reflect all trends. So, for example, it is inappropriate to compare flights from the USA and from the UAE for one simple reason: the Emirates airline mentioned above uses very large aircraft [16]: Airbus A380 (119 aircraft, the largest operator of this type) and Boeing 777 with a passenger capacity of 509 and 386 people, respectively, and the airline Delta Air Lines uses [17] mainly Boeing 737, Airbus A321 with a passenger capacity of 189 and 200 people, respectively, that is, in order to transport the same number of passengers, Emirates will need to make 2.5 times fewer flights than Delta Air Lines, due to which such connections will be weaker on visualization.

The task of forecasting aircraft production

To solve the problem of forecasting aircraft production, a model from the Facebook company Prophet was chosen. Since this is a time series forecasting task, the choice was not very large, but this model enjoys authority due to a large number of advantages compared to other models: it can automatically detect annual, monthly and daily seasonality, it supports cyclicality, it can make predictions for long periods ahead, it is easy configurable, works great with non-stationary series, learns quickly and makes predictions.

With a data set that contains information about all the aircraft loaded, we can filter the data by a certain condition and perform an analysis: Airbus A320 family aircraft production statistics, 4-engine aircraft production statistics and piston engine aircraft production statistics. To begin with, let's depict these statistics on graphs (Fig. 9-11).
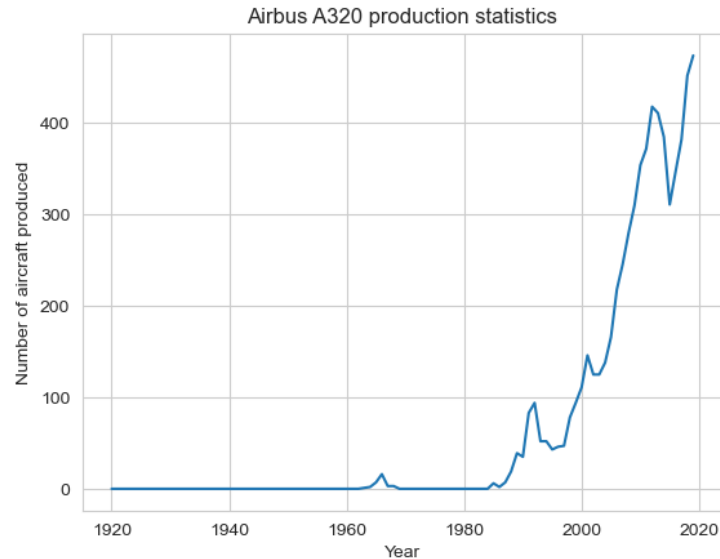

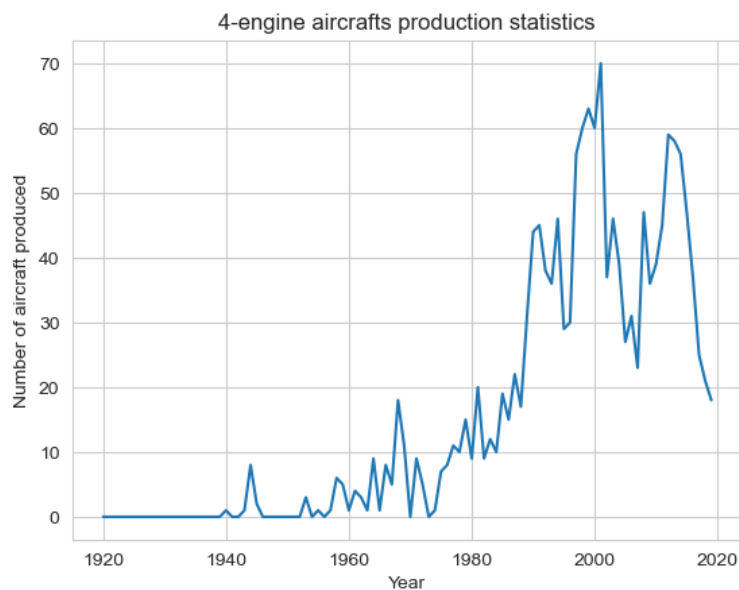*Fig. 9.* The number of Airbus A320 model aircraft built by year

Fig. 10. The number of built aircraft with 4 engines by year

After decomposing these data sets into trend, seasonality, and residuals, we see that all these data sets lack a clear trend and seasonality. Fortunately, the Prophet model can still make good predictions. For the Prophet model, the most important parameter is changepoint_prior_scale, which indicates how strongly the model will be affected by new

values. For each data set, they are unique, so cross-validation was used to select the ideal parameter value. We will build models with it, and we can also display forecasts together with confidence intervals (Fig. 12-14).

Fig. 11. The number of built aircraft with piston engines by year
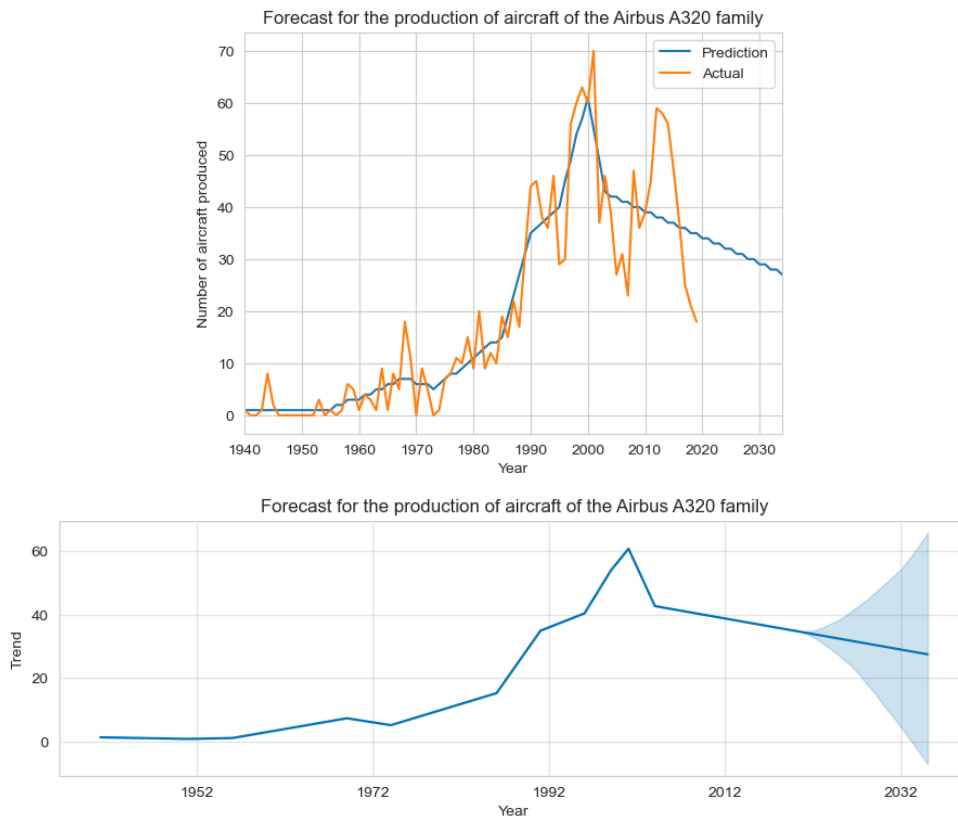

Fig. 12. Predictions of the model regarding the production of Airbus A320 aircraft
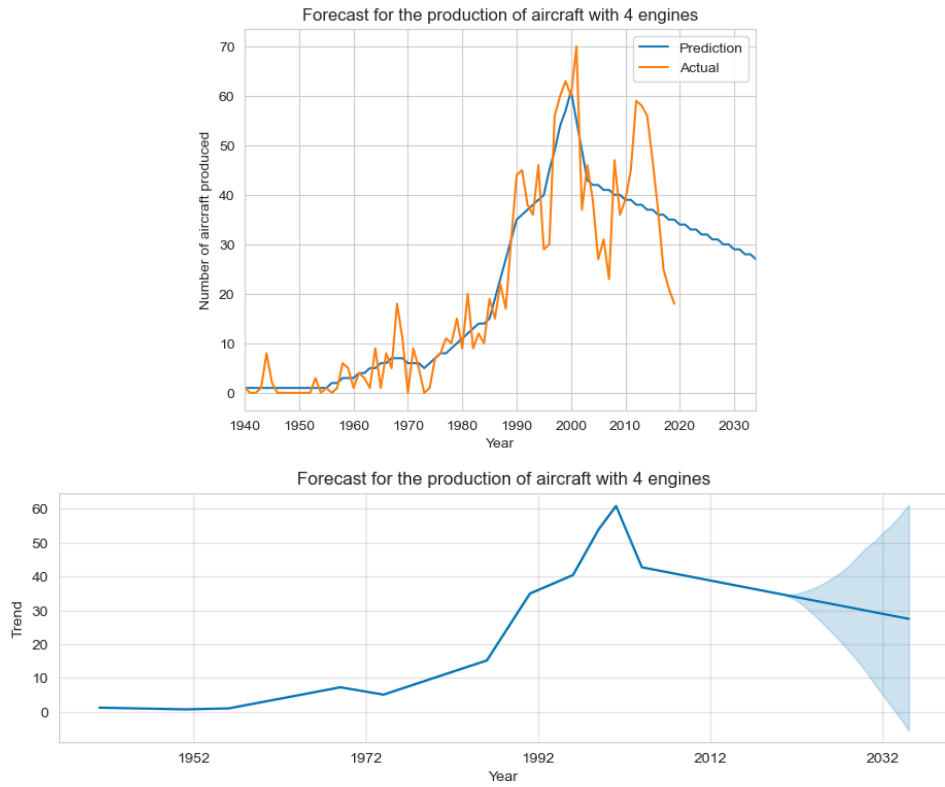
Fig. 13. Predictions of the model for the production of aircraft with 4 engines
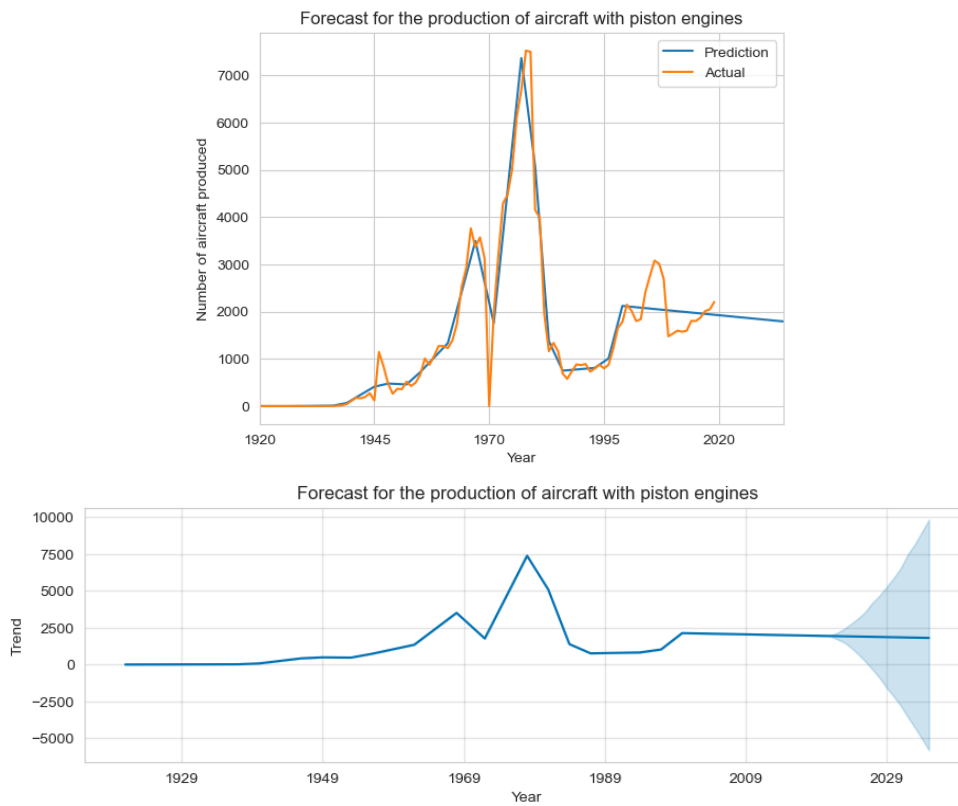

Fig. 14. Predictions of the model regarding the production of aircraft with piston engines

We can see that the model predicts a further increase in the production of Airbus A320 aircraft, a gradual decline in the production of 4-engine aircraft, and a more or less constant production of piston-engine aircraft. In general, these predictions coincide with banal logic, because the Airbus A320 plane is now on the rise: it is used by an incredible number of airlines, 4-engine planes are becoming a thing of the past, and piston engines have lived, live, and will live because this is the main type of engine for small planes that are very popular abroad.

In summary, it can be noted that quite high accuracy was obtained during cross-validation, the model repeats the training data well, but is not overtrained, since it does not repeat the training data completely, which allows us to treat these predictions with sufficient confidence. However, it is also worth noting that the confidence intervals for the 4-engine and piston-engine aircraft production forecasts are quite wide, indicating some uncertainty in the model's long-range predictions.

Conclusion

To divide countries into clusters based on the number of flights between them, a clustering model was created using the Louvain method, after which three visualizations were generated: with all connections, with the three most important connections for each country, and with the hundred most important connections in the world. During the clustering analysis, 5 clusters were identified, which mainly depended on the geographical location of the countries. Nevertheless, it was observed that Ukraine has the most flights with Russia in the last 3 years, although there is no direct connection at all. This is caused by the fact that Crimean airports belong to Ukraine in the downloaded data sets.

Such a model can be applied even outside of aviation because it indicates not only the geographical proximity of certain countries but also the geopolitical and economic situation in these countries, as well as the relations between different countries.

Airlines, in turn, can use such visualizations to analyze which countries can be "captured" and which already have a very serious player on the market, and the competition there will be very difficult.

A model was built to forecast the production of Airbus A320, 4-engine and piston-engined aircraft. The changepoint_prior_scale hyperparameter was selected using cross-validation. The resulting models produced the following result: production of Airbus A320 aircraft would increase, 4-engine aircraft would decrease, and piston-engine aircraft would remain more or less constant.

These models will provide aircraft manufacturers with solid knowledge about which aircraft to build and which to buy for airlines to increase profits through the production of trendy aircraft on the one hand and reduce losses due to fleet obsolescence on the other.

In general, the results of the conducted analysis open up new opportunities for the aviation industry, however, it is worth noting that further research and development in this

direction can bring even more useful information and open new horizons for the progress of the aviation industry.

REFERENCES

1. Airbus A380: End of a multibillion-dollar dream. URL: https://www.dw.com/en/airbus-a380-the-end-of-a-multibillion-dollar-dream/a-60124995 (from: 20.06.2023).

2. Keras: Depp Learning for humans. URL: https://keras.io (from: 29.05.2023).

3. Louvain Community Detection. URL: https://github.com/taynaud/python-louvain (from: 29.05.2023).

4. Prophet | Forecasting at scale. URL: https://facebook.github.io/prophet (from: 29.05.2023).

5. Gephi – The Open Graph Viz Platform. URL: https://gephi.org (from: 29.05.2023).

6. nginx. URL: https://nginx.org (from: 29.05.2023).

7. *Dube, K., Nhamo, G., Chikodzi, D.* COVID-19 pandemic and prospects for recovery of the global aviation industry // Journal of air transport management, 2021, Vol. 92. DOI: 10.1016/j.jairtraman.2021.102022

8. *Pereira, B.A., Lohmann, G., Houghton, L.* Innovation and value creation in the context of aviation: a Systematic Literature Review // Journal of air transport management, 2021, Vol. 93. DOI: 10.1016/j.jairtraman.2021.102076

9. *Cui, Qiang; Yu, Li-Ting* A Review of Data Envelopment Analysis in Airline Efficiency: State of the Art and Prospects // Journal of advanced transportation, 2021, Vol. 2021. DOI: 10.1155/2021/2931734

10. *Tarasonok D. Y., Oliinyk Y. O., Likhouzova T. A.* Models for forecasting flight delays // Inter-branch scientific and technological digest «Adaptive systems of automatic control» № 2(43), 2023

11. Aircraft Type Designators. URL: https://www.icao.int/publications/doc8643/pages/search.aspx (from: 29.05.2023).

12. Open data @ OurAirports. URL: https://ourairports.com/data (from: 29.05.2023).

13. Fast unfolding of communities in large networks. URL: https://arxiv.org/pdf/0803.0476.pdf (from: 29.05.2023).

14. Reingold Layout. URL: https://www.sciencedirect.com/topics/computer-science/reingold-layout (from: 29.05.2023).

15. Кластеризація. URL: https://flights.compich.com/clustering (from: 29.05.2023).

16. Emirates Fleet Details and History. URL: https://www.planespotters.net/airline/Emirates (from: 29.05.2023).

17. Delta Air Lines Fleet Details and History. URL: https://www.planespotters.net/airline/Delta-Air-Lines (from: 29.05.2023).