Applied Computer Science, vol. 19, no. 4, pp. 51–67
doi: 10.35784/acs-2023-35

Submitted: 2023-08-23 | Revised: 2023-09-18 | Accepted: 2023-09-22

Keywords: weather prediction, Facebook Prophet, aviation, machine learning, data analysis

Sylwester KORGA [0000-0002-1327-0320]*, *Kamil ŻYŁA* [0000-0002-6291-003X]*, *Jerzy JÓZWIK* [0000-0002-8845-0764]**, *Jarosław PYTKA* [0000-0002-5474-3585]***, *Kamil CYBUL* ****

PREDICTIVE TOOLS AS PART OF DECISSION AIDING PROCESSES AT THE AIRPORT – THE CASE OF FACEBOOK PROPHET LIBRARY

Abstract

Prophet is a quite fresh and promising open-source library for machine learning, developed by Facebook, that gains some significant interest. It could be used for predicting time series taking into account holidays and seasonality effects. Its possible applications and deficit of scientific works concerning its usage within decision processes convinced the authors to state the research question, if the Prophet library could provide reliable prediction to support decision-making processes at the airport. The case of Radawiec airport (located near Lublin, Poland) was chosen. Official measurement data (from the last 4 years) published by the Polish Government Institute was used to train the neural network and predict daily averages of wind speed, temperature, pressure, relative humidity and rainfall totals during the day and night. It was revealed that most of the predicted data points were within the acceptance threshold, and computations were fast and highly automated. However, the authors believe that the Prophet library is not particularly useful for airport decision-making processes because the way it handles additional regressors and susceptibility to unexpected phenomena negatively affects the reliability of prediction results.

1. INTRODUCTION

Decision-making processes at the airport are crucial for ensuring safety and efficient airport operations. The personnel responsible for making decisions must take into account various factors such as weather conditions, air traffic load, aircraft conditions, as well as regulations and procedures. Decisions made at the airport have a direct impact on flight

* Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Department of Computer Science, Poland, s.korga@pollub.pl, k.zyla@pollub.pl
** Lublin University of Technology, Faculty of Mechanical Engineering, Department of Production Engineering, Poland, j.jozwik@pollub.pl
*** Lublin University of Technology, Faculty of Mechanical Engineering, Department of Motor Vehicles, Lublin, Poland, j.pytka@pollub.pl
**** Lublin University of Technology, Doctoral School of the Lublin University of Technology, Mechanical engineering, k.cybul@pollub.pl; Central Office of Measures, Warsaw, Poland, kamil.cybul@gum.gov.pl

operations, passenger movement, and the safety of all individuals present at the airport. Their inaccuracy might cause serious hazards like plane crashes, emergency landings, or simply unnecessarily increases the risk of pilots' activities in the air or airport staff on the ground.

Decision making activities at the airport involve very complicated and time consuming processes requiring specialized skills and knowledge. A very significant part of it is weather prediction, that influences wide range of topics from possible crisis management, to decisions regarding take-offs and landings of aircrafts.

Modern applicable decision-making tools use advanced prediction methods (Dee et al., 2011; El Hachimi et al., 2021), which are highly sophisticated in nature and require large amounts of data and computing power, as well as human mathematical and engineering skills. However, the accuracy of predictions is still limited to the complexity of the processes taking place, which negatively affects the reliability of decisions.

While seeking for better solutions, computational models for predicting particular phenomena were employed. Development of such models is associated with data exploration and analysis. The common problem is the lack (or low level) of match between forecast and actual conditions. Not to mention the challenges of managing large data sets, which is another class of problems.

These cause people to search for even better methods and tools aiding decision making processes. One of the methods is machine learning, when computer algorithms try to predict or recognize some phenomena based on the already gathered data. Historical data, which is represented as time series, might be troublesome for the traditional machine learning algorithms, especially when dealing with fast changes in time of interconnected indicators.

One of machine learning libraries, that is specifically designed for handling time series, is Facebook prophet (Prophet, 2023). The library utilizes methods of forecasting data series based on the additive model, in which non-linear trends are aligned (adjusted) to the annual, weekly and daily seasonality, as well as holiday effects. The way the Prophet is handling time series and promising results that are provided as an outcome, convinced the authors to take a closer look at this library in context of the weather prediction for the purpose of aiding decision making processes at the airport.

2. FACEBOOK PROPHET LIBRARY

Prophet is an open-source library for machine learning developed by Facebook (Krieger, 2021; Prophet, 2023). Machine learning was defined in the literature ages ago as a method of acquiring knowledge by computer systems based on gathered data (Mitchell, 1997). There are many libraries available for this purpose, such as TensorFlow (2023) and Keras (2023). Nevertheless, Facebook Prophet is the central point of this article as a promising new library focusing the community interest. Also the scientific literature concerning this library usage is very limited. Section 2.1 covers the literature overview on the Prophet library application. Section 2.2 gives introduction to the internals of the library.

2.1. Prophet library applications

In order to find the Facebook Prophet library usages, the authors searched through Scopus and Web of Science databases. These databases were chosen as major sources of high quality scientific works. In both cases:

- the issued query was “Facebook AND Prophet”,
- all available databases were included in the search,
- title, abstract and keywords underwent the search,
- no date constraint was applied.

Scopus returned 108 results with the date span 2015-2022. Only 9 results were related to weather prediction. Web of Science returned 64 results with the date span 2015-2022. Only 6 results were related to weather prediction. In total, 10 distinct articles related to weather prediction were found. No article was found on weather prediction for airport decision-making.

After analyzing abstracts of the works that were found, it was revealed that the Prophet library finds applications in a big variety of human life domains. First field is electrical engineering, e.g. to predict anomalies in photovoltaic components (Bendiek et al., 2022), or predict electric current usage (Banga et al., 2021). Another one is Internet of Things, e.g. monitoring safety (by detecting anomalies) of the environment around children (Shenbagalakshmi & Jaya, 2022) or anomaly detection in sewer air temperature sensor system (Thiyagarajan et al., 2020). When it comes to a social perspective, Prophet is used e.g. to analyze the change of public opinion trend from a quantitative perspective (Qiu et al., 2019) or to forecast the success of television series (Akula et al., 2019). In the case of economics it was used e.g. for cash flow (Weytjens et al., 2021) and stock price prediction (Garlapati et al., 2021). At last, Prophet is used in the domain of medicine e.g. to predict disease spread (Patil & Pandya, 2021) or hospitalization rate (Ryu et al., 2021).

From this article point of view, the most important domain of the Prophet library applications is weather prediction. Haq (2022) assessed the impact of temperature and rainfall values in all Himalayan states on water resources planning and management. The Prophet model was implemented to forecast and assess the performance of the Climate Deep Long Short-Term Memory (CDLSTM). According to the author, Prophet showed good performance and accuracy.

Junsuk and Tae (2021) used Prophet for meteorological data (e.g. temperature, humidity, and wind speed) forecasting that could help estimate the probability of occurrence and spreading speed of wildfires. The Prophet model was trained using meteorological data between 2016 and 2018 from Goseong, Gangwon-do. It was effective in terms of computing speed and predicting the overall trend. However, prediction of sudden irregular changes was unsatisfactory. This work was written in Korean, thus this summary follows the claims from the English abstract only (the authors do not speak Korean).

Toharudin et al., (2021) used Prophet to predict air temperature only, in Bandung. It was compared with the Long Short-Term Memory (LSTM) method. According to the authors, Prophet performed slightly better on maximum air temperature prediction, while LSTM was slightly better on minimum air temperature prediction.

Asha et al., (2020) used Prophet to predict air temperature only, in Kerala India, in order to allow some planning activities on governmental level to protect peoples’ health and agriculture fields against heat. It was compared with the random forest algorithm. According to the authors, both solutions gave comparable results, but random forest was slightly more accurate.

Oo and Phyu (2019) described a system for monitoring the weather in a greenhouse where environmental data are continuously monitored by sensors. Prophet was used to predict the temperature only. Results were validated by comparing the forecasted temperature with the actual value. Authors concluded that the validation confirmed Prophet potential for successful application to microclimate prediction in sustainable agriculture.

Papacharalampous et al., (2018) investigated the predictability of monthly temperature and precipitation by applying automatic univariate time series forecasting methods (naive, random walk, AutoRegressive Fractionally Integrated Moving Average (ARFIMA), exponential smoothing state-space model with Box–Cox transformation, ARMA errors, Trend and Seasonal components (BATS), simple exponential smoothing, Theta, Prophet) to a sample of 985 40-year-long monthly temperature and 1552 40-year-long monthly precipitation time series. Results of predictions were compared to the real measurement data. Authors concluded that Prophet is competitive to the other methods, especially when it is combined with externally applied classical seasonal decomposition, and that it could be used in long-term applications.

Sulasikin et al., (2021) used Seasonal AutoRegressive Integrated Moving Average (SARIMA), LSTM and Prophet methods to predict the monthly rainfall in the next 24 months in central Jakarta, which might be useful for the purpose of a data-driven policy on how to face potential flood. Results were compared to the real measurement data. Authors concluded that Facebook Prophet, with the lowest Mean Squared Error and Root Mean Squared Error, is the fittest model to predict the monthly rainfall in the Central Jakarta.

Soltaganov et al., (2018) used Prophet to create an additive regression model based on a time series of the monthly precipitation sums measured on the Tomsk station (station synoptic index 29430) between 1996 and 2016. It was compared to a SARIMA model. Authors concluded that in the case of stationary time series the SARIMA prediction model is more accurate.

El Hachimi et al., (2021) investigated several machine learning models to create a decision support platform with the aim of making agriculture in Morocco more efficient and sustainable. Prophet was used for the forecasting of the hourly average air temperature only. It was compared with LSTM, which was then evaluated as less accurate.

Narmeen et al., (2022) used Prophet to forecast the spread (number of cases) of COVID-19 in major Pakistan cities depending on the weather. Authors focused on the impact of temperature only. No clear conclusion about the technical capabilities of Prophet was made.

2.2. Prophet library mechanics

Prophet is a library designed for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality. Holiday effects are taken into account as well. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well (Krieger, 2021).

Time series analysis has two goals - first, detecting the nature of a phenomenon which is represented by a sequence of observations; second - predicting future values of a time series. Prophet should seamlessly match the weather data, which is a time series, as each data record consists of timestamp and measurements of particular weather indicators, such as temperature, air pressure, wind speed, etc. Prophet is based on the Fourier's series, in order to provide an elastic model of seasonal effects useful in analysis of time series. The library automatically detects changes in data trends and chooses data points where the change occurred (Krieger, 2021).

Prophet implements functions from generalized linear models (GLM), as well as additive models (AM), mainly by extending GLMs with nonlinear smoothing functions. It was described by Taylor and Letham (2018). The main difference between Prophet and other statistical methods/tools is the approach to analysis, that allows an analyst to apply his domain knowledge in a prediction algorithm, without knowing the exact internals of statistical methods inside the algorithm. Such an approach supports processes and methods of prediction based on domain experts' decisions.

The Prophet library can be applied to simple and complex time series that include single or multiple seasonality, holidays, and data trends. Seasonality might follow different patterns, like days, weeks, years. According to (Taylor & Letham, 2018), the library uses a decomposable time series model

$$Y(t) = g(t) + s(t) + h(t) + \varepsilon t \quad (1)$$

where: $g(t)$ – piecewise linear or logistic growth curve for modeling non-periodic changes in time series,
$s(t)$ – periodic changes (e.g. weekly and yearly seasonality),
$h(t)$ – effects of holidays (provided by a user) with irregular schedules,
εt – error term, accounts for any unusual changes not accommodated by the model.

Whatmore, Prophet allows to adjust accordingly to the needs the following parameters (Prophet, 2023):

1. Changepoints - points where trends change. They can be found by the Prophet algorithms or they can be set up by an analyst.
2. Seasonality - periodic functions that can influence time series. By default, Prophet takes into account yearly, weekly and daily seasonality, and tries to find trends that represent these periodic effects in data.
3. Holidays - special days (holidays or other recurring events) that can be modeled by a Prophet's additive model. Dates can be set up manually, or built in the prediction environment.
4. Fourier's order - pace of seasonality function changes. It influences the degree of model fit (there is a risk of a model over-fitting).

Prophet is a library that uses time as a regressor when automatically fitting linear and nonlinear functions. Modeling of seasonality as an additive component is the same approach as the one applied in the case of Holt-Winters exponential smoothing. As a result, the problem of prediction is treated as an exercise with curve fitting, instead of looking for the direct dependence on the time of each observation in the time series (Prophet, 2022; Prophet, 2023).

3. RESEARCH GOAL

Prophet is a quite fresh and promising open-source library for machine learning developed by Facebook (Krieger, 2021; Prophet, 2023), that gains some significant interest. To mention, 172 works in total were indexed in Scopus and Web of Science between 2015-2022. Unfortunately, based on the query described in Section 2, only 10 works concern weather at all, and no works concern the topic of predicting the weather for the purpose of aiding decision processes at the airport. Moreover, no works explore the full complexity of weather prediction based on multiple weather components, like wind, temperature, pressure, humidity, etc., that could be useful for the purpose of such decision making.

There are few more reasons why the topic of this article is focused on the Prophet library. It was designed to deal with time series with seasonality. It should smoothly fit the weather data, which is a typical time series, as each data record consists of a timestamp and measurements of particular weather components. Finally, it is recommended by some specialists in the topic as a tool for weather forecasting, according to personal interviews conducted by this article authors.

The abovementioned factors convinced the authors to state the following research question: Could the Prophet library provide reliable weather prediction data for the purpose of aiding decision processes at the airport? The authors would like to acknowledge that finding the best weather forecasting model is intentionally out of scope of this work.

4. MATERIALS AND METHODS

In order to answer the stated research question, the authors used real publicly accessible meteorological data gathered by the professional synoptic station, located in Radawiec at a small airport. The following subsections describe the origin of data, tools and procedure that led to results and conclusions.

4.1. Synoptic station in Radawiec

Real weather data used in this article was gathered by the synoptic station in Radawiec near the city of Lublin. The station is managed by IMGW (Institute of Meteorology and Water Management – National Research Institute, Warsaw, Poland) and it was assigned the number of 351220495. It was chosen as the result of local patriotism of the authors.

Fig. 1. Satellite view of the synoptic station (red pin) in Radawiec and the LALR airport (blue pin)

The station is located at the airport in Radawiec Duży 272A, 21-030 Motycz, Poland - see Fig. 1. The airport name is Radawiec Aeroclub Airport (LALR). Coordinates of the airport and the synoptic station are WGS-84: 51°13'18" N, 22º23'40" E. Distance from the synoptic station to the airport runway is approximately 500 m.

4.2. Procedure of meteorological data acquisition

The synoptic station in Radawiec belongs to IMGW and performs measurements in full accordance with this national institute guidelines and requirements. Everything comply with regulations on the international level as well (IMGW, 2022; Urząd Lotnictwa Cywilnego, 2022).

Meteorological data is collected by a so-called meteorological cage, which is a specially designed box containing meteorological instruments performing measurements. The box is made of white painted wood and has a shutter structure, which allows for free exchange of air inside it and minimizes absorption of sunlight. It is placed on a special stand, so that the instruments inside are not directly illuminated by the sun and remain 2 m above the ground. Access to the inside is possible thanks to the door opening to the North. The standard equipment of the meteorological cage includes a psychrometer (humidity measurement), maximum thermometer, minimum thermometer, thermograph and a hygrograph (for measuring and recording changes in relative air humidity as a function of time). Cloudiness is measured on the octant scale from 0 to 8 (IMGW, 2022).

A synoptic station works (registers data) in a continuous manner. The time at all synoptic stations managed by IMGW is synchronized - universal time allows to refer to the same moment in time. A single set of measurements is made during the first 5 minutes of each hour. Registered data is sent to the IMGW server, and then shared in the public domain (IMGW, 2022).

4.3. Software and hardware

The most important software components, that were used in order to obtain the research goal, are as follows:

1. Anakonda - version 3, Python language interpreter, Jupiter IDE.
2. Pystan - version 2.19.1.1, additional files facilitating Facebook Prophet installation.
3. FBProphet - version 0.7.1, Facebook Prophet library for weather prediction.
4. Pandas - version 1.1.3, data management tool.
5. Bokeh - version 2.2.3, data visualization.
6. R - version 3.6.2, statistical computations.

Specialized libraries matplotlib and datetime were used to generate the charts, and aliases were assigned to them - as shown in Fig. 2.

```
In [36]: import matplotlib.pyplot as plt
         import datetime as dt

         prediction
         model.plot(prediction)
         left = dt.date(2021, 1, 1)
         right = dt.date(2021, 3, 31)
         plt.gca().set_xbound(left, right)

         plt.plot(Fin_2021.year, Fin_2021.ADT, 'rx--',)
         left = dt.date(2021, 1, 1)
         right = dt.date(2021, 3, 31)
         plt.gca().set_xbound(left, right)


         plt.show
```

Fig. 2. A snippet of the program listing illustrating the invocation of the prediction process and generating charts for the analyzed data

Computations were performed using a laptop: Dell 5540 Intel Core i7-9750H, RAM 16 GB, SSD 512 GB. It was controlled by Windows 10 operating system, which was updated to the latest version.

4.4. Data structure and cleaning

Meteorological data from the Radawiec synoptic station, used for the purpose of this experiment, are certified data that were downloaded from the IMGW website. Data were available in CSV-type files that contained daily average values of particular weather indicators (IMGW, 2022).

Unfortunately downloaded data were not sufficiently structured, thus it required additional curation before providing it to Prophet. This paragraph mentions some preprocessing actions. First, comma was used as a separator of a value in a file row, as well as a separator of a fraction part in numbers. In case of numbers, comma has been replaced with dot. Second, the coding of files has to be switched to ISO-8859-1. Another problem were statuses "8" (measurement missing) and "9" (phenomenon missing) defined by IMGW - they were missing in the data records. In the case of something missing IMGW just put zero there. What's more, each date was split into 3 columns - year, month, day. For the purpose of computation, they had to be joined into one value of datetime type.

Pandas library was used to assign column headers to the data, as shown in Fig. 3. After the process of segregating columnar data, a vertical concatenation was performed for the analyzed files - as shown in Fig. 4.

```
In [ ]: df2018['year']=pd.to_datetime(df2018[['year', 'month', 'day']])
        df2019['year']=pd.to_datetime(df2019[['year', 'month', 'day']])
        df2020['year']=pd.to_datetime(df2020[['year', 'month', 'day']])

In [ ]: df2021_01['year']=pd.to_datetime(df2021_01[['year', 'month', 'day']])
        df2021_02['year']=pd.to_datetime(df2021_02[['year', 'month', 'day']])
        df2021_03['year']=pd.to_datetime(df2021_03[['year', 'month', 'day']])
```

Fig. 3. Column headers encoding for data from CSV files

```
In [ ]: Fin = pd.concat([df2018, df2019, df2020], axis= 0, ignore_index=True)

In [ ]: Fin_2021=pd.concat([df2021_01, df2021_02,df2021_03], axis= 0, ignore_index=True)
```

Fig. 4. A snippet of the program listing used for vertical concatenation of files

Tab. 1. Data structure containing parameters included in the research process

Column name	Measurement SI Unit	Description by IMGW	Exemplary record
station_code	N/A	Code of a synoptic station	351220495
station_name	N/A	Name of a synoptic station	Lublin – Radawiec
date	N/A	Date of a value measurement	2011-01-01
ADWS	km/h	Average daily wind speed	6.9
ADT	ºC	Average daily temperature	-0.7
MDP	hPa	Mean daily pressure	1008.5
ADRH	%	Average daily relative humidity	89.8
TPDD	mm	Total precipitation during the day	978.6
TPDN	mm	Total precipitation during the night	1008.5

The files with data did not have the captions - they have to be added by the authors based on the IMGW guidelines and descriptions. Table 1 explains structure of the processed data records. The authors skipped columns that were unnecessary from the Prophet computation and the article's goal point of view. N/A was put in case of no SI unit associated.

The weather data used for the purpose of the experiment concerned years 2018-2021. The total number of data records was 1158 (this number includes leap years) - each day was represented by 1 data record. 1096 records were a training set and 62 records were used to verify predictions. All data records were complete - no empty values were present, no data records were rejected due to their validity.

4.5. Procedure

The procedure of the experiment involved the following steps:

1. Downloading real measurement data of weather indicators for Radawiec synoptic station.
2. Preprocessing of the acquired data to fit the requirements of the Prophet library functions.
3. Neural network training.
4. Prediction of weather indicators for the Radawiec airport.
5. Making projection of the real data on the predicted values.

The real measurements data from 01-01-2018 to 31-12-2020 were used for the purpose of network training. A set of predictions for weather indicators was made for the next 62 days starting from 01-01-2021. Then the predicted values were compared with the real values measured by the station.

Each of the weather indicators was predicted independently of other weather indicators. The Prophet library by default does not take into account the correlation of many weather indicators. No holidays were set up, data from all days of a week were taken into account. Seasonality was turned on, as the weather repeats a more or less similar pattern each year. Process of neural network training was automated then by the Prophet library. This process was shown in Fig. 5.

```
In [28]: from fbprophet import Prophet
         model = Prophet()

In [29]: model.fit(baza1)

         INFO:fbprophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.

Out[29]: <fbprophet.forecaster.Prophet at 0x18b41407070>

In [31]: future_dates = model.make_future_dataframe(periods=365)

In [32]: prediction = model.predict(future_dates)
```

Fig. 5. A snippet of the program listing calling the model training process

The authors decided to predict the most important weather indicators, being: average daily wind speed, average daily temperature, mean daily pressure, average daily relative humidity, total precipitation during the day, total precipitation during the night. Each of the indicators has assigned an acceptance rate, which is a percentage of predicted values that fit within the uncertainty band computed by the Prophet library.

Moreover, descriptive statistics was computed based on the difference between the real and predicted data - minimum, maximum, median, mean, standard deviation, box plots. The authors calculated as well: Median Absolute Percent Error (MAPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the maximum residual error (MAX). Such approach can be found e.g. in (Ortiz-Bejar et al., 2020; Papacharalampous et al., 2018). These metrics give a chance to compare results presented in this work to results obtained by other researchers. The R environment and Metrics library were used to make calculations.

The MAX metric captures the worst-case error between the predicted value and the actual value, as expressed by

$$MAX(x, xp) = max|x_i - xp_i| \tag{2}$$

where: x – vector of real measurement data,
xp – vector of predicted data.

The MAE metric measures the accuracy of a model fit, as expressed by

$$MAE = (1/n) * \Sigma|x_i - xp_i| \tag{3}$$

where: x – vector of real measurement data,
xp – vector of predicted data,
n – total number of observations,
Σ – sum operation.

RMSE metric measures how far predicted values are from observed values in a regression analysis. In other words, it shows the data are concentrated around the line of the best fit. It is expressed by

$$RMSE = \sqrt{}[\Sigma(xp_i - x_i)^2 / n] \tag{4}$$

where: x – vector of real measurement data,
xp – vector of predicted data,
n – total number of observations,
Σ – sum operation,
$\sqrt{}$ – square root operation.

MAPE metric measures the forecasting accuracy of a model. MAPE was described by Ortiz-Bejar et al., (2020) as easy to interpret, scale-independent, and one of the most widely used to measure regression systems performance. It is expressed by

$$MAPE = (1 / n) * \Sigma(|x_i - xp_i| / |x_i|) * 100 \tag{5}$$

where: x – vector of real measurement data,
xp – vector of predicted data,
n – total number of observations,
Σ – sum operation.

5. RESULTS

Data used to train the network and make the prediction took 508 kB of disk space. The network training and prediction time was 2 minutes 45 seconds in total.

Table 2 presents the weather indicators that underwent the prediction based on the trained network model. Abbreviations from the table are used in the following sections of this article. Acceptance rates might be perceived as being around 80 % in most cases. The worst rates were obtained for ADT (56 %) and TPDD (61 %). Such a big difference might be the result of some unusual meteorological phenomena. The Prophet library calculates predictions based on few years of historical data and gives results that could be interpreted in the following way: "if the weather is repeatable, then you should expect the predicted results". Lack of reliance on actual meteorological conditions makes the library vulnerable to unexpected events, which can be seen in the case of ADT and TPDD.

Tab. 2. Acceptance rates for weather indicators undergoing the prediction

Abbreviation	Full name	Acceptance
ADWS [km/h]	Average daily wind speed	85 %
ADT [ºC]	Average daily temperature	56 %
MDP [hPa]	Mean daily pressure	80 %
ADRH [%]	Average daily relative humidity	80 %
TPDD [mm]	Total precipitation during the day	61 %
TPDN [mm]	Total precipitation during the night	88 %

Fig. 6-11 present prediction charts for each of the weather indicators. The purpose of such charts is to show how the prediction performed. Dark blue line is the line of predicted data trend. Bright blue area around the line determines the upper and lower confidence interval - the width of the area is determined by the Prophet library on its own. Red line presents how the real measurement data points (values from the control set - 62 days starting from 01-01-2021) fit the prediction. It can be seen that the worst quality of prediction was obtained for ADT, and next for ADWS. In other cases all or almost all real measurement data points were inside the confidence interval.

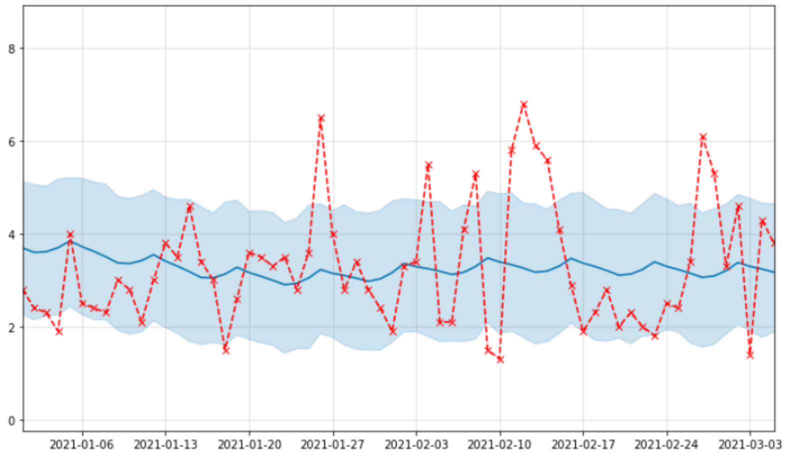

Fig. 6. Prediction chart for average daily wind speed (ADWS) in km/h

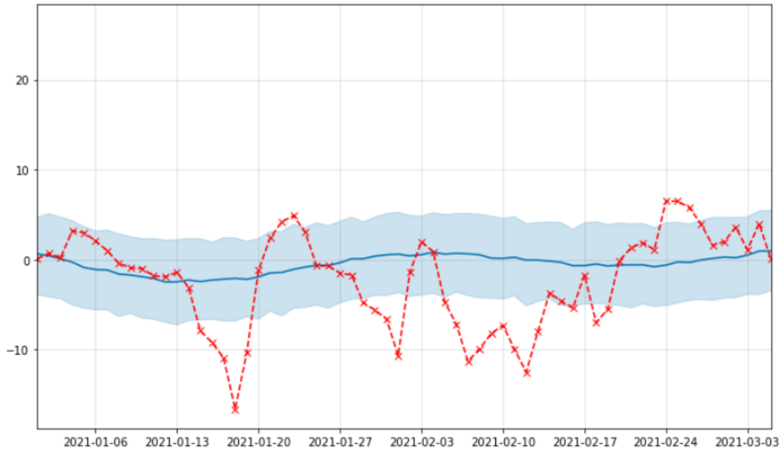

Fig. 7. Prediction chart for average daily temperature (ADT) in ºC

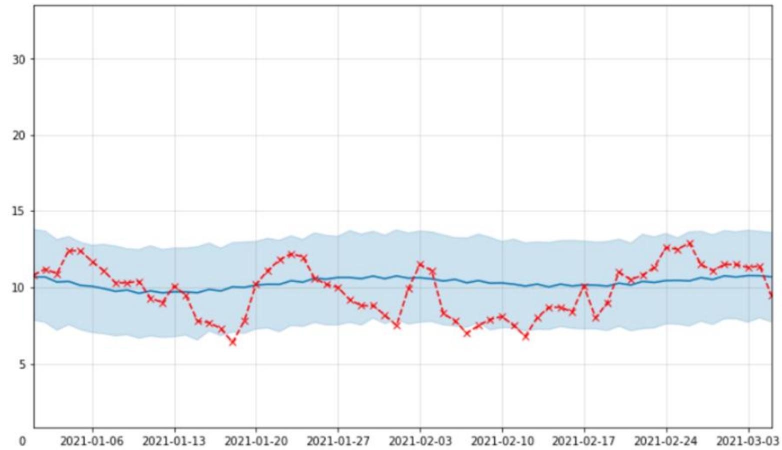

Fig. 8. Prediction chart for mean daily pressure (MDP) in100x hPa

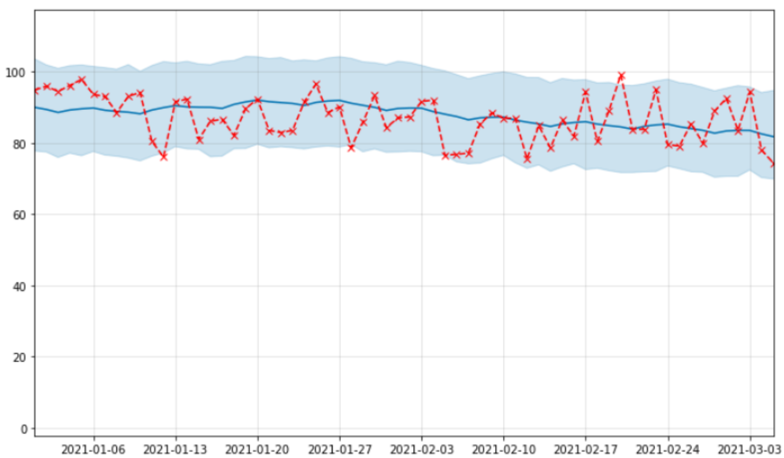

Fig. 9. Prediction chart for average daily relative humidity (ADHR) in %

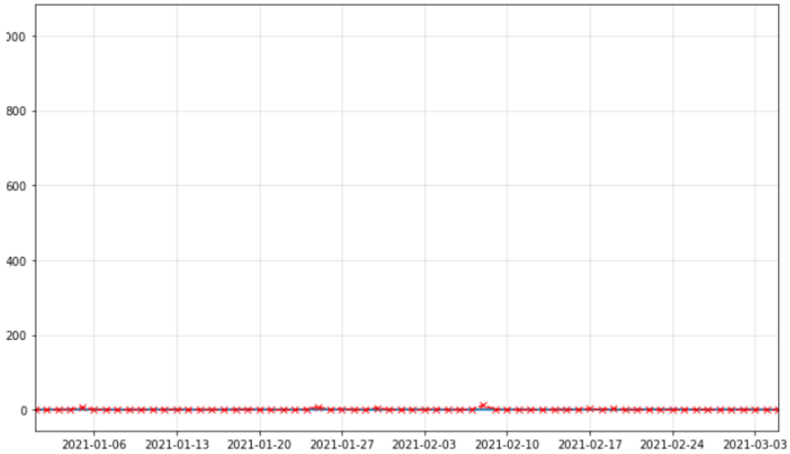

Fig. 10. Prediction chart for total precipitation during the day (TPDD) in mm

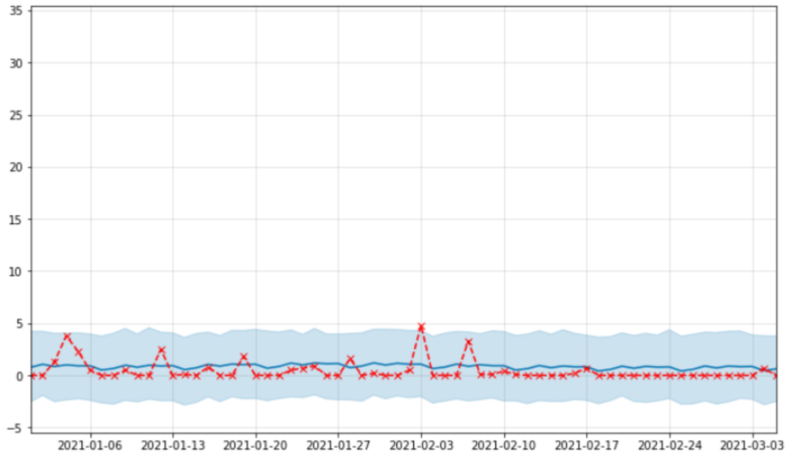

Fig. 11. Prediction chart for total precipitation during the night (TPDN) in mm

Table 3 presents the basic statistics for predicted values and how they differ from actual measurements originating from the synoptic station. Predicted weather indicators are in columns, and in rows are the following statistics for delta values: minimum, 1st quartile, median, mean, standard deviation, 3rd quartile, maximum. Figure 12 shows box plots based on the delta values. By delta value the authors understand an absolute value from the difference between measured value of a weather indicator for the particular day and predicted value for the same day.

Computed values confirm the worst outcomes for ADT. Large values can be seen associated with ADHR, although humidity is measured in the highest range of values (0-100 %) compared to the other weather indicators. In terms of MAPE the best result was achieved by ADHR and MDP being respectively 0.07 and 0.27. In case of RMSE and MAE the best results were achieved by TPDN (1.02 and 0.59 respectively), ADWS (1.36 and 1.07 respectively) and MDP (1.70 and 1.41 respectively).

Tab. 3. Statistics computed for the predicted data.

	ADHR [%]	ADT [ºC]	ADWS [km/h]	MDP [hPa]	TPDD [mm]	TPDN [mm]
minimum	0.20	0.06	0.04	0.00	0.04	0.00
1st quartile	2.55	1.13	0.47	0.55	0.23	0.17
median	5.37	3.62	0.87	1.35	0.50	0.37
mean	5.72	4.25	1.07	1.41	1.07	0.59
standard deviation	3.91	3.60	0.84	0.96	2.07	0.84
3rd quartile	7.83	6.72	1.36	2.13	0.74	0.53
maximum	17.45	14.37	3.64	3.58	12.70	4.34
MAX	17.45	14.37	3.64	3.58	12.70	4.34
MAE	5.72	4.25	1.07	1.41	1.07	0.59
RMSE	6.91	5.55	1.36	1.70	2.32	1.02
MAPE	0.07	21.21	0.34	0.27	4.87	3.58

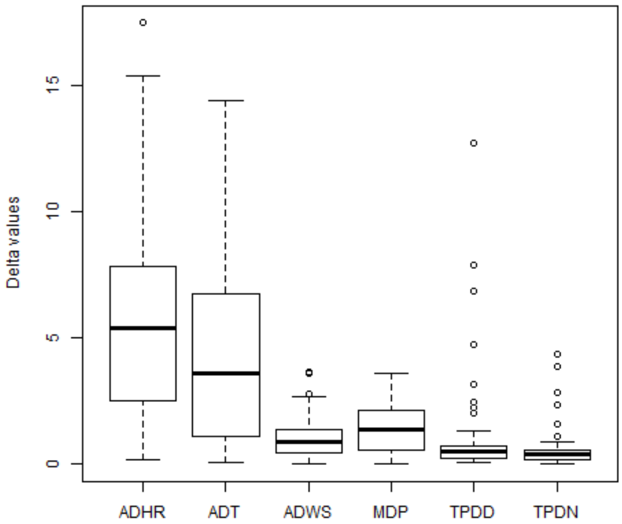

Fig. 12. Box plots showing delta values distribution for particular weather indicators

6. DISCUSSION AND CONCLUSIONS

Prophet is an emerging library provided by Facebook, that could be used for predicting time series taking into account holidays and seasonality effects. The authors stated the research question, if the Prophet library could provide reliable weather prediction data for the purpose of aiding decision processes at the airport. The case of Radawiec airport was chosen. Predictions lasting 62 days were performed for the following weather indicators: average daily wind speed, average daily temperature, mean daily pressure, average daily relative humidity, total precipitation during the day, total precipitation during the night. No other authors performed such analysis to such an extent.

Average daily values of meteorological data were used for the prediction in the article. The authors understand that such data are less useful for aviation purposes than average hourly data. Although the authors wanted to base their work on real-life data from a government agency, and daily data from the IMGW institute were broadly accessible for the public - thanks to that our experiment is more replicable. Public accessibility of the data was one of the factors considered when choosing the dataset.

Due to the chaotic nature of meteorological processes, predicting the weather based on historical data might be inaccurate. Weather patterns that occurred in the previous years might not occur in the next year. Such a thing took place in the case of data used for the purpose of this article. First, the authors fed the neural network with 10 years long data, although there was a slight climate shift during that time. That caused the 10-years model to be a worse basis for the prediction than the 3-years model used in the article. What is more, something unexpected (not following the predicted pattern) occurred in the weather, and it can be especially well seen in the case of an average daily temperature and an average daily wind speed. The authors find these as one of the biggest flaws of the Prophet library in the context of decision making at the airport.

Each of the weather indicators was predicted independently of other weather indicators. The Prophet library by default does not take into account the correlation of many weather indicators. For example the speed of wind is predicted by default based solely on the vector with the wind data. Introducing correlation of the wind with other weather indicators, like air pressure, temperature, etc. is named as adding new regressors (Prophet, 2022). What is surprising, the values of the regressor for the time of prediction (e.g. 62 days, like in this article) have to be known before the prediction. If they are not known, then they have to be predicted in some other way, before they could be used as a regressor. To give an example: someone wants to forecast the daily temperature in the next month, but depending on the average daily pressure, which is not known at that point as well. Someone would expect that both values would be predicted while automatically maintaining mutual influence of these weather indicators. The authors find the above-described facts, and the procedure of adding regressors, as another flaw of the library, which might be an interesting topic for further research as well.

To sum up, process of neural network training is highly automated by the Prophet library. The training and prediction is quite fast, even for medium-size datasets and a personal computer. Although, on the other hand, the library users have a very narrow influence on the procedure of obtaining the final network model. Moreover, using additional regressors is burdensome and prediction outcomes are vulnerable for unexpected weather phenomena. Thus the authors conclude that the Prophet library is not suited for providing reliable weather prediction data for the purpose of aiding decision processes at the airport. Rather, some other solutions should be pursued.

Funding

This research was funded by Lublin University of Technology.

This work was prepared within the project PM/SP/0065/2021/1 entitled "Method for airplane takeoff and landing distance measurement", financed by the Ministry of Education and Science (Poland) as a part of the Polish Metrology Programme.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

Akula, R., Wieselthier, Z., Martin, L., & Garibay, I. (2019). Forecasting the success of television series using machine learning. *ArXiv, abs/1910.12589*. https://doi.org/10.48550/arXiv.1910.12589

Asha, J., Rishidas, S., SanthoshKumar, S., & Reena, P. (2020). Analysis of temperature prediction using random forest and Facebook Prophet algorithms. *Lecture Notes on Data Engineering and Communications Technologies, 46*, 432-439. https://doi.org/10.1007/978-3-030-38040-3_49

Banga, A., Ahuja, R., & Sharma, S. C. (2021). Stacking machine learning models to forecast hourly and daily electricity consumption of household using Internet of Things. *Journal Of Scientific & Industrial Research*, *80,* 894-904.

Bendiek, P., Taha, A., Abbasi, Q. H., & Barakat, B. (2022). Solar irradiance forecasting using a data-driven algorithm and contextual optimisation. *Applied Sciences*, *12*(1), 134. https://doi.org/10.3390/app12010134

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., … Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

El Hachimi, C., Belaqziz, S., Khabba, S., & Chehbouni, A. (2021). Towards precision agriculture in Morocco: A machine learning approach for recommending crops and forecasting weather. *2021 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA)* (pp. 88–95). IEEE. https://doi.org/10.1109/ICDATA52997.2021.00026

Garlapati, A., Krishna, D. R., Garlapati, K., Srikara Yaswanth, N. M., Rahul, U., & Narayanan, G. (2021). Stock Price Prediction Using Facebook Prophet and Arima Models. *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1–7). IEEE. https://doi.org/10.1109/I2CT51068.2021.9418057

Haq, M. A. (2022). CDLSTM: A novel model for climate change forecasting. *Computers Materials & Continua, 71*(2), 2363-2381. https://doi.org/10.32604/cmc.2022.023059

IMGW. (2022). *Homepage of the Institute of Meteorology and Water Management - National Research Institute.* https://www.imgw.pl

Junsuk, K., & Tae, J. K. (2021). Application of Facebook's Prophet model for forecasting meteorological data. *Journal of the Korean Society of Hazard Mitigation, 21*(2), 53-58. https://doi.org/10.9798/KOSHAM.2021.21.2.53

Keras. (2023). Keras library homepage. https://keras.io

Krieger, M. (2021, February 20). Time series analysis with Facebook Prophet: How it works and how to use it. https://towardsdatascience.com

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill Science.

Narmeen, M., Sattar, M. U., Khan, H. W., Fatima, M., Azad, M.-D., & Ghani, F. (2022). Impact of Weather on COVID-19 in Metropolitan Cities ofPakistan: A Data-Driven Approach. *International Journal of Computing and Digital Systems*, *11*(1), 905–915. https://doi.org/10.12785/ijcds/110174

Oo, Z. Z., & Phyu, S. (2019). Microclimate Prediction Using Cloud Centric Model Based on IoT Technology for Sustainable Agriculture. *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)* (pp. 660–663). IEEE. https://doi.org/10.1109/CCOMS.2019.8821705

Ortiz-Bejar, J., Ortiz-Bejar, J., Zamora-Mendez, A., Pineda-Garcia, G., Graff, M., & Tellez, E. S. (2020). Forecasting without context problem. *2020 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (pp. 1–6). IEEE. https://doi.org/10.1109/ROPEC50909.2020.9258744

Papacharalampous, G., Tyralis, H., & Koutsoyiannis, D. (2018). Predictability of monthly temperature and precipitation using automatic time series forecasting methods. *Acta Geophysica, 66,* 807-831. https://doi.org/10.1007/s11600-018-0120-7

Patil, S., & Pandya, S. (2021). Forecasting dengue hotspots associated with variation in meteorological parameters using regression and time series models. *Frontiers in Public Health, 9*, 798034. https://doi.org/10.3389/fpubh.2021.798034

Prophet. (2022). Seasonality, holiday effects, and regressors. https://facebook.github.io/prophet/docs/seasonality,_holiday_effects,_and_regressors.html

Prophet. (2023). Facebook Prophet library homepage. https://facebook.github.io/prophet/

Qiu, Q., He, J., Chen, H., & Qiu, J. (2019). Research on the evolution law of emergency network public opinion. *2019 12th International Symposium on Computational Intelligence and Design (ISCID)* (pp. 157–161). IEEE. https://doi.org/10.1109/ISCID.2019.10119

Ryu, S., Nam, H. J., Kim, J. M., & Kim, S. W. (2021). Current and future trends in hospital utilization of patients with schizophrenia in Korea: A time series analysis using national health insurance data. *Psychiatry Investigation, 18*(8), 795-800. https://doi.org/10.30773/pi.2021.0071

Shenbagalakshmi, V., & Jaya, T. (2022). Application of machine learning and IoT to enable child safety at home environment. *Journal of Supercomputing, 78,* 10357–10384. https://doi.org/10.1007/s11227-022-04310-z

Soltaganov, N. A., Sherstnev, V. S., Sherstneva, A. I., Botygin, I. A., & Krutikov, V. A. (2018). Construction of predictive models of meteorological parameters of the atmospheric surface layer. *IOP Conference Series: Earth and Environmental Science, 211*, 012027. https://doi.org/10.1088/1755-1315/211/1/012027

Sulasikin, A., Nugraha, Y., Kanggrawan, J. I., & Suherman, A. L. (2021). Monthly rainfall prediction using the Facebook Prophet model for flood mitigation in central Jakarta. *2021 International Conference on ICT for Smart Society (ICISS)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICISS53185.2021.9532507

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45. https://doi.org/10.1080/00031305.2017.1380080

TensorFlow. (2023). TensorFlow library homepage. https://www.tensorflow.org

Thiyagarajan, K., Kodagoda, S., Ulapane, N., & Prasad, M. (2020). A temporal forecasting driven approach ysing facebook's prophet method for anomaly detection in sewer air temperature sensor system. *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 25–30). IEEE. https://doi.org/10.1109/ICIEA48937.2020.9248142

Toharudin, T., Pontoh, R. S., Caraka, R. E., Zahroh, S., Lee, Y., & Chen, R. C. (2021). Employing long short-term memory and Facebook prophet model in air temperature forecasting. *Communications in Statistics-Simulation and Computation*, 52(2), 279-290. https://doi.org/10.1080/03610918.2020.1854302

Urząd Lotnictwa Cywilnego. (2022). *Meteorological service for international air navigation. Annex 3 to the Convention on International Civil Aviation*. https://www.ulc.gov.pl/_download/prawo/prawo_miedzynarodowe/konwencje/Zal%C4%85cznik_3_cz_I_cz_II.pdf

Weytjens, H., Lohmann, E., & Kleinsteuber, M. (2021). Cash flow prediction: MLP and LSTM compared to ARIMA and Prophet. *Electronic Commerce Research, 21,* 371-391. https://doi.org/10.1007/s10660-019-09362-7