ISM
INTERNATIONAL
SCHOOL OF MANAGEMENT
University of Applied Sciences

Hausarbeit / assignment

Sommersemester 2021

X Erstversuch / first attempt
O 1. Wiederholung / second attempt
O 2. Wiederholung / third attempt

Studiengang / study program	Business Intelligence and Data Science
Modul/Fach / module/subject	Business Control
Thema / topic	Evaluation of analytical software on atmospheric carbon dioxide data
Dozent / lecturer	Joakim Nägele
Student (Nachname, Name / surname, first name)	Jagadish, Prajwal Neuß, Annika Haid, Philipp
Abgabetermin / deadline for submission	31.08.2021

Bewertung des Dozenten / lecturer's grading:

Punkte / points	
/ Erreichte (reached) / von maximal (of maximum)	
Datum / date	
Unterschrift / signature	
Punkte gesamt / points total	**/ 100** **Erreichte (reached) / von max. 100 (of max. 100)**
Unterschrift / signature	

EVALUATION OF ANALYTICAL SOFTWARE ON ATMOSPHERIC CARBON DIOXIDE DATA

Paper

ISM
INTERNATIONAL SCHOOL OF MANAGEMENT
University of Applied Sciences

Philipp Haid
philipp.haid.2020@student.ism.de

Annika Neuß
annika.neuss.2020@student.ism.de

Prajwal Jagadish
prajwal.jagadish.2020@student.ism.de

Business Intelligence & Data Science
Business Control
Joakim Nägele

Hamburg,
August 31st, 2021

Abstract

The carbon dioxide emission has risen dramatically the past years. Climate change has experienced and increasing importance over the past years. To understand the extent of the climate change and to analyze the predictable of the carbon dioxide emission this paper evaluates different analytical software on atmospheric carbon dioxide data. This paper focusses on two research questions:

(I) Which analytical software has a high forecasting performance for atmospheric carbon dioxide data?
(II) Which analytical software has the most intuitive process of implementing carbon dioxide forecasts?

For the evaluation three analytical software have been chosen: Tableau, PowerBI and IBM Cognos. The analytical software were chosen because of its different position on the 2021 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms, which evaluates different analytical software by ability to execute and completeness of vision.

In order to profoundly answer research question (I) a forecasting period of 20 years was created and moreover, different Key Performance Indicators were defined. During the forecasting horizon the performance of the analytical software change. The analysis shows that the RMSE of Tableau has been steady throughout the period of time whereas PowerBI's score has been in the beginning rather low but then increased rapidly. Similar behavior was indicated with the MAPE, with the difference that IBM Cognos has the steadiest performance. Nevertheless, there are more measures needed to clearly state the best performance.

Moreover, the software have been tested according to their intuitive process of implementing carbon dioxide forecasts. Therefore, the implementation process of each analytical software was created. As a result, different process steps and functionalities were discovered in the tools.

Content

List of figures

List of tables

1. Introduction

Arising environmental issues due to ever increasing energy utilization have led to specific interest in gas emission forecasting. A multitude of environmental problems have become a global concern with the major contribution of greenhouse gases. These greenhouse gases add up to only a small percentage of our atmosphere, but disproportionately impact the earth's radiant heat and warm our planet which is commonly known as the greenhouse effect. All these greenhouse gases account for a smaller portion including methane of the whole atmosphere, which include nitrous oxide, ozone, and greenhouse carbon dioxide. A drastic increase of the temperature also raises the atmospheric humidity which leads to heating our planet in a vicious cycle. Forecasting carbon dioxide emissions is essential for climate policy decisions. Forecasts of carbon dioxide emissions help not only to predict future global warming but also estimate expected costs of emission reductions and expected benefits from preventing global warming. This is also a key element which also helps the public to be more conscious about environmental issues. The scientific reviews and amount of literature on forecasting carbon dioxide through different software are considerably less discussed and reviewed. Forecasting of carbon dioxide (CO2) is impactful in many sectors such as agriculture to reduce the production of CO2. Technologies to capture and store CO2 help in maintaining the wellness of our environment. Evaluating the performance of each analytical tool is essential in the whole process as scientific discussions are limited by these tools. There are different methods used in forecasting CO2, for example, artificial intelligence, simulation forecasting, optimal growth models, and so on. The major concern of this paper is which analytical tool should be used to forecast the performance of CO2 emissions.

1.1 Objectives

The main objective of the paper is to analyze and implement: (i) which analytical software has a high performance regarding forecasting CO2 emissions and (ii) which software has the most intuitive process of implementing CO2 forecasts. The aim and purpose of this research is to have a streamlined workflow in the analytical tool which reduces the work of other researchers and scientists by automating organized tasks and by implementing advanced software development designs. It is our mission to provide efficient modeling and simulation frameworks with verified behavior to help the reproduction of complex environmental issues. This enables the effective understanding of the scenario and behavior of the predictive system. The focus of this paper is to use different analytical tools such as business intelligence platforms and forecasting methods to produce relevant CO2 predictions based on attributes like carbon trend analysis, seasonal patterns through exponential smoothing and relevant Time indicators. This process will also use linear forecasting in different visualization tools to produce adequate CO2 predictions based on data features.

1.2 Structure

In our approach to the paper, we have demarcated it into four different sections. The initial section is about describing problems and objectives with insights into structure and delimitation. The following section is about the literature review which describes the forecasting process and defines how each tool and application are applied with respect to forecasting a model on different software using diverse sources. The third section is about research methodology which presents the methodological approach taken in this paper and justifies the data acquired and analytical tools used. The next section is about the findings where different methods of forecasting are implemented with different evaluation methods according to the prescribed data set. This section describes where different analytical tools are applied to forecast or predict the data. However, it is compared with other tools to evaluate the performance of each tool, which impacts our whole process by knowing which analytical software performs the best in forecasting or predicting atmospheric carbon dioxide. The last section is a discussion which gives a broad view of the overarching ideas, and this is followed by the conclusion.

1.3 Delimitation

The purpose of the research is to propose a simplified implementation of forecasting methods and analytical tools through different visualization techniques. The analytical tools used in the process of forecasting are Tableau, Microsoft power-BI and IBM Cognos analytics. These three powerful tools are used to evaluate the performance. Related requirements and a growing number of standards regarding tools or applications have made the process simpler. In this paper, we have come across various techniques to evaluate the model which are Seasonality, trend analysis, MASE, MAE, MAPE, and RMSE. To handle the time series data, triple exponential smoothing has been used to equate trend, stationary and seasonal. Regarding the different descriptions of the model, we evaluate performance based on the above techniques for a 20-year forecasting horizon. The main emphasis in visualization will be seasonality or trend analysis which makes better understanding in forecast or prediction. The resulting visualized patterns describe which tool or software has better performance in forecasting CO2 emissions. The visualization technique makes the process faster and helps in the decision-making processes related to performance.

Apart from the methods used in this paper, it is further possible to extend the model by adding machine learning algorithms like artificial neural networks, optimal growth models, or grey models. Forecasting CO2 emissions could be done by dividing two different categories in data mining models. One is a descriptive model that extracts information about the data set, and the other one a predictive model that estimates the target value (dependent variable). The use of the gray prediction method estimates the usage of carbon emissions in a region and identifies energy consumption that has a major effect on the atmosphere. With regard to other methods, there are three models that are compared to estimate Co2 emissions, and those models are ARIMA, Grey model, and non-linear gray Bernoulli model (NGCM-OP).

2. Literature Review

The following literature offers theoretical insights on the theory behind. Furthermore, the applied model and the error measurements for future evaluation are explained.

2.1 Definition

Forecasting is the process of estimating or predicting based on present or historic data, the analysis of forecasting is commonly based on trends. A forecast is successful, if the result explains various parameters like trend analysis, seasonality, or exponential smoothing. The forecasts of CO2 emissions are made to analyze the prediction and provide effective insights to implement better decisions for companies and to solve environmental issues.

Forecasting can be nebulously defined into different categories based on the period over which it is done. The distinctions largely are dependent on which causal factors are assertive at what forecasting horizons. Hence any methods of forecasting used should take these operating factors into consideration.[1]

- For instance, in electric load forecasting, short term forecasts will be dominated by weather (driving air conditioning / heating use), whereas long term forecasts will be dominated by economic development and political decisions (building offshore wind parks will yield different load distributions than power plants in the middle of the country).

Forecasting is categorized into two different methods:

(i) Quantitative methods: To forecast demand, Quantitative methods use mathematical models that rely on historical data, utilizing large amounts of data and figures for effective integration.

(ii) Qualitative methods: This approach uses factors based on the expert's knowledge, experience, intuition, value systems and so on.

Therefore, forecasting regulates the comparison, evaluating with actual measures and reviewing for modifications to ensure smooth predictions.[2]

2.2 Goals

The iterative approach in predicting or forecasting is dependent on how well the planning process is implemented in defining a goal. Our goal is to evaluate performance with different analytical tools by implementing the model available on the platform. The evaluation procedure is implemented with visualization tools like Tableau, PowerBI, and Cognos. In these tools, the performance is measured by relative squared error, mean absolute scaled error, root mean squared error, mean absolute percentage error and seasonal time frame. These error values define how well the analytical tools perform in forecasting carbon dioxide. The comparison of all the performance fulfills our mission through which analytical

[1] (Kolassa, 2014)
[2] (Box, 2013)

tools can be implemented with regard to emissions of CO2. The diagram below explains the process of forecasting. Planning is essential to predict acceptable outcomes. However, the best plan can be evaluated and implemented if the entire planning process is monitored.[3]

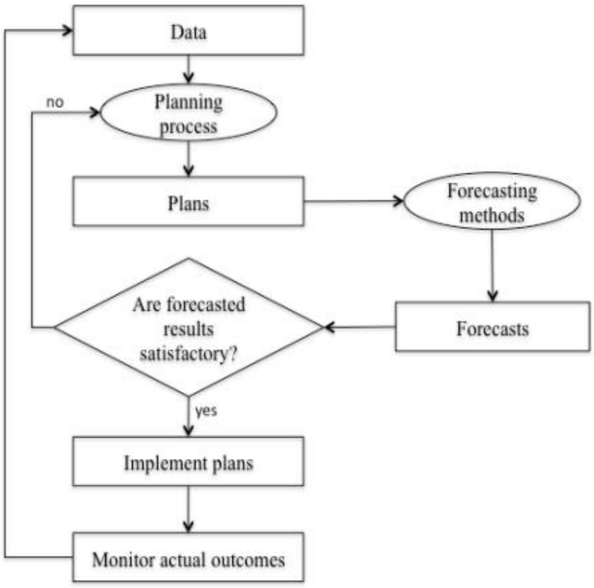


Figure 1: Forecasting process[4]

2.3 Methods and applications

To forecast carbon dioxide there are many different analytical tools to analyze and implement. In the process of analytics and business intelligence, data visualization plays an important role in taking raw data and delivering efficient information analysis that makes interpretation and understanding easier. The key elements in applying different applications for analytics and business intelligence platforms are based on augmented capabilities that are easy-to-use functionality and effective analytical workflow. The impact of decision-making in data visualization lies in cloud ecosystems which support building, deploying, and managing analytics. With respect to analytics and business intelligence platforms, data visualization is no longer differentiated. This platform has included integrated support for enterprise reporting and augmented analytics (machine learning and artificial intelligence).

The forecasting functions are drawn from statistical analysis of historic information used in different forecasting methods. Many variables impact the accuracy of the forecast. Forecasting function guides the measures of future values that have to be predicted. The following are types of forecasting methods:

---

[3] (Armstrong, Green, & Graefe, 2015)
[4] (Armstrong, Green, & Graefe, 2015)

- Trend (linear or straight line): In this method, forecasting technique is based on linear regression of time series forecasting. This gives the best forecasting reliability when major factors influence in measuring forecast.

  The formula for trend forecasting is

  $y = at + b$

  where y is the dependent variable (for example, revenue), t is the independent time variable,

$$a = \frac{N\left(\sum_{i=1}^{N} t_i y_i\right) - \left(\sum_{i=1}^{N} t_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N\left(\sum_{i=1}^{N} t_i^2\right) - \left(\sum_{i=1}^{N} t_i\right)^2} \quad \text{(the slope of the trend line)}$$

  and

$$b = \frac{\left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} t_i^2\right) - \left(\sum_{i=1}^{N} t_i\right)\left(\sum_{i=1}^{N} t_i y_i\right)}{N\left(\sum_{i=1}^{N} t_i^2\right) - \left(\sum_{i=1}^{N} t_i\right)^2} \quad \text{(the intercept)}$$

- Growth (curved or curved line): Exponential regression technique of time series is used in growth forecasting methods. This gives the best forecasting reliability when major factors influence in measuring exponentially.

  The formula for Growth forecasting is

  $y = ba^t$

- Autoregression (seasonal): The auto-correlation function is used in autoregression forecasting. This forecasting technique detects linear, nonlinear, and seasonal fluctuations from historic data that tends to forecast trends into the future.[5]

  The formula for Autoregression forecasting is

$$y_t = \sum_{j=1}^{M} d_j y_{t-j}$$

where

[5] (IBM, 2014)

$$\sum_{j=1}^{M} \phi_{j-k} d_j = \phi_k \qquad (k = 1, \ldots, M) \qquad (d_j \text{ are the linear prediction (LP) coefficients})$$

and

$$\phi_j = \langle y_i y_{i+j} \rangle \approx \frac{1}{N-j} \sum_{i=1}^{N-j} y_i y_{i+j} \qquad \text{(auto-correlation of the historic series)}$$

.

The smoothing models attempt to smooth the fluctuations in time series by smoothing or averaging. Exponential smoothing is one of the statistical methods commonly used for forecasting.

Exponential smoothing: In this forecasting method weighted averages of past observations are taken to produce forecasts. In this framework forecasts are generated quickly and for a wide range of time series that have great influence and leading application for industries.

Triple exponential smoothing is an extension of exponential smoothing that supports seasonality. This method is also known as Holt-Winters exponential smoothing. In this method the hyperparameter are: alpha, beta, gamma, phi (damping coefficient) and seasonality type is additive or multiplicative.[6]

$$S_t = \alpha(y_t/I_{t-L}) + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad \mathit{OVERALL\ SMOOTHING}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \qquad \mathit{TREND\ SMOOTHING}$$

$$I_t = \beta(y_t/S_t) + (1-\beta)I_{t-L} \qquad \mathit{SEASONAL\ SMOOTHING}$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \qquad \mathit{FORECAST}$$

In this paper, the analytical tools are measured by their performance and evaluation of particular models. So, to evaluate performance and accuracy of the forecast, error methods are implemented to analyze the forecast.

There are different types of error used to evaluate accuracy of the model. The following explains types of error used to evaluate forecasting carbon dioxide:

- Mean absolute percentage error (MAPE): MAPE is the average of the percentage errors. This measure is most commonly used to evaluate forecast accuracy. MAPE divides each error individually by demand, high errors during low-demand periods will have a significant impact on MAPE.

[6] (Hyndman & Athanasopolos, 2018)

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

- Mean Absolute error (MAE): MAE measures the average absolute difference between the actual and predicted numeric output. MAE does not change the dimensionality of the errors and is also less sensitive to outliers.

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n}.$$

- Mean squared error (MSE): This method is faster to compute and simple to manipulate when compared to RMSE. But, with regard to original error this method is not scaled and that becomes difficult to measure in statistical forecasting models.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

- Root Mean squared error: In predicting quantitative data this method is standardly used. RMSE is the square root of the average of squared errors.[7]

2.4 Tools

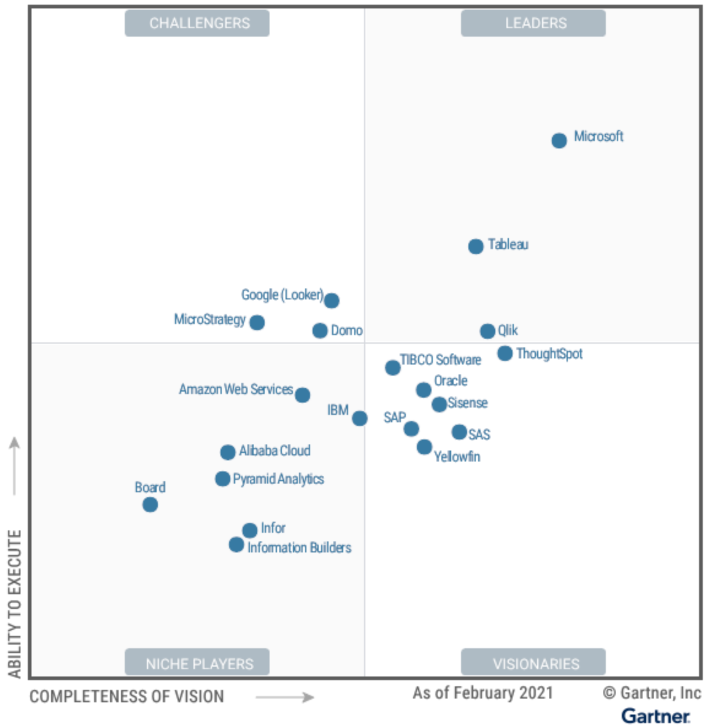


Figure 2: Magic Quadrant for Analytics and Business Intelligence Platforms[8]

[7] (NIST / SEMATECH, 2012) & (Shcherbakov, et al., 2013)
[8] (Gartner, 2021)

Figure 2 describes different analytical applications which are used in analytical and business intelligence platforms. The graph explains two different parameters. The x-axis describes the completeness of vision (performance of each tool), niche players (small segments), visionaries (group of successful long-term leaders), and the y-axis explains the execution process, how well each tool is evaluated in the execution phase.[9]

Analytical and business intelligence platforms have created a standard process with regard to dashboard, reporting and integrated process. This implementation of different analytical tools is effective and efficient in choosing the right analytical platform.

From Gartners Quadrant three tools haven been chosen: Tableau, Microsoft PowerBI and IBM Cognos. Thereby the focus was on the two market leaders. Moreover, to better identify strength and weaknesses of the analytical software the third tool, IBM Cognos was chosen from a rather lower position compared to PowerBI and Tableau.

Tableau

Tableau is a powerful analytical tool that enables expert product capabilities and expands its product offering with regard to augmented analytics and governance capabilities. To enhance the access, preparation, analysis, and presentation of data, Tableau takes a visual-based exploration approach. This accords Tableau its prominent position in the Magic quadrant. The improved augmented analytics and governance capabilities make this a formidable service. Tableau is price effective and has a push when users choose subscription pricing. In augmented analytics tableau imported Ask data and Explain data to offer automated insights and natural language query. Tableau is also essential in three crucial areas of Cloud, artificial intelligence, and embedded analytics. However, Tableau is the leader of visual exploration and data manipulation with regard to customer's requirement and choosing this platform is a right choice with all the aspects related to Forecasting CO2.

Microsoft PowerBI

Microsoft PowerBI is a leader in business intelligence and analytics platforms. This tool offers various features like data preparation, interactive dashboards, visual-based data discovery, and augmented analytics. This tool can also be used as a stand-alone and free personal analysis tool. Microsoft PowerBI provides end –to- end solutions with regard to business intelligence platforms. In PowerBI report server SaaS option is available that runs in Azure cloud or on-premises option. The users are not given the flexibility to prefer a cloud IaaS in Microsoft PowerBI but it is provided in Azure.   In business intelligence platforms this tool is always mentioned by users and to integrate with them. So, this platform is powerful when compared with other analytical tools.

[9] (Kolassa, 2014) & (Richardson, Schlegel, Sallam, Kronz, & Sun, 2021)

IBM Cognos

IBM Cognos analytics offers visual exploration, enterprise reporting, governed and augmented analytics in a single platform. IBM Cognos analytics plays an important role in augmented analytics with defining statistical methods with time series forecasting and detection. This application is for the existing Cognos customers that have developed to build and modernize. Cognos analytics are highly priced when compared to other large cloud providers. IBM offers customer requirements with different deployment features and major infrastructure as a service through its own license. Therefore, capabilities of all the above tools mentioned for evaluation are standard with respect to forecasting carbon dioxide.

3. Research Methodology

In order to profoundly answer the research questions:

(I) Which analytics software has a high forecasting performance for atmospheric carbon dioxide data?

(II) Which analytical software has the most intuitive process of implementing carbon dioxide forecasts?

different research methods have been used.

3.1 Dataset

The chosen the dataset "Trends in Atmospheric Carbon Dioxide"[10] shows the values of carbon dioxide in the years 1958 to 2018 and was selected because it includes clean data points and therefore does not need any further data cleansing. This enables the primarily focus on the use of the data. Furthermore, the topic of carbon dioxide emission and the carbon footprint has become more important the past years and thus also has a current relevance. The data are sourced from the US Government's Earth System Research Laboratory, Global Monitoring Division.[11] In total 607 data points have been collected. Table 1 describes the data.

column name	**description**
Average	• The monthly mean CO2 mole fraction determined from daily averages. • If there are missing days concentrated either early or late in the month, the monthly mean is corrected to the middle of the month using the average seasonal cycle. • Missing months are denoted by -99.99.

Interpolated	• Values from the average column and interpolated values where data are missing.
	• Interpolated values are computed in two steps.
	  ○ First, for each month the average seasonal cycle in a 7-year window around each monthly value is computed. In this way the seasonal cycle is allowed to change slowly over time.
	  ○ Second, the trend value for each month by removing the seasonal cycle is determined; this result is shown in the trend column.
	• Trend values are linearly interpolated for missing months.
	• The interpolated monthly mean is then the sum of the average seasonal cycle value and the trend value for the missing month.

[10] (Data Hub, 2018)
[11] (Global Monitoring Labatory, 2021)

Trend	• seasonally corrected
Number of days	• -1 denotes no data for number of daily averages in the month.

Table 1: Description of the CO2 dataset[12]

3.2 Forecasting methods

A forecasting method is especially needed to answer research question (I). Therefore, this research question is based on quantitative research because it follows statistical, mathematical, and computational techniques.[13] Since the time series data is containing a seasonal component, triple exponential smoothing is used as a forecasting method. It takes trend as well as seasonality into account by using three equations and three constants. Trend is defined as the increasing or decreasing value in the series whereas seasonality is the repeating of the short-term cycle in the series. Both trend and seasonality can be multiplicative or additive. A value of a time series variable is considered as the resultant of the combined impact of its components. The components of a time series get either multiplicated (multiplicative model) or added (additive model). Furthermore, the additive model does not vary in amplitude and frequency over time, but the multiplicative model does. Moreover, there are different smoothing factors: alpha for data, beta for trend and gamma is the seasonal smoothing factor. All range from zero to one and must be estimated in such a way that the error is minimized.[14] This forecasting has been not only chosen because of the seasonal component but also because it is one of the simplest, quickest, and most common method to be implemented. Therefore, all analytical software should be able to implement this method.

3.3 Evaluation methods

In order to make well-founded statements about which software is most suitable to forecast carbon dioxide data there are different metrics to measure the best performance. The two key performance indicators that were chosen to evaluate the performance are MAPE and RMSE. The **Mean Absolute Percentage Error** (MAPE) is one of the most commonly used KPI.[15] It gets calculated by the sum of the individual absolute errors divided by the demand of each period separately. The **Root Mean Squared Error** (RMSE) is defined as the square root of the average squared error. It is commonly used in climatology, forecasting and regression analysis and therefore suitable for the prediction of carbon dioxide data. If the RMSE is small, it indicates that the predictions are close to the true responses. Furthermore, a value of zero would connote a perfect fit to the data. In comparison to the MAPE the RMSE does not treat each error equally but rather puts more value on the most significant errors. This results in a poor score if one big error is detected. Through these two different

[12] (Data Hub, 2018)
[13] (Conroy, Kaye, & Schantz, 2008)
[14] (Jain & Malehorn, 2005, pp. 334-335)
[15] (Vandeput, 2019)

error assessments the performance can be compared in a wider range because it not only focusses on one evaluation parameter.

With respect to research question (II), which analytical software has the most intuitive process of implementing carbon dioxide forecasts, a qualitative research methodology has been taken in comparison to the previous research question. The intuition of implementing the forecast has been rated through the analysis of the implementation process of each tool. Additionally, to the intuitive interface the forecasting possibilities as well as guidance through the software were important evaluation factors in order to give a profound statement to the second research question. Especially for this research question the choice of analytical software was important. Since the focus was on an intuitive process it was important, that there has not been any using knowledge of the tools. Because knowing the steps of the implementation would manipulate the approach of intuition, which is defined as the immediate cognition without the use of conscious rational processes.[16]

[16] (Leach & Weick, 2017)

4. Findings

When implementing forecasts, Tableau and Cognos provide the user with details about the used model. Since each tool uses exponential smoothing as its forecasting model, information about various coefficients is provided and the model is specified. Tableau and Cognos also calculate the performance of the specific model based on the training set. PowerBI does not provide further details about the used model. Table 2 summarizes the details of the different models used in the forecasts.

	Cognos	**Tableau**	**PowerBI**
Trend	Additive	Multiplicative	-
Seasonality	Multiplicative	Multiplicative	-
MASE	0,225	0,22	-
MAE	0,239	0,3	-
RMSE	0,307	0,3	-
MAPE	0,10%	0,10%	-
Seasonal timeframe	12	12	-
Alpha	0,533	0,5	-
Beta	0,014	0,039	-
Gamma	0,234	0,108	-

Table 2: Comparison of forecasting models

While the performance of the two models is similar, there is a difference in the coefficients. In addition, the Cognos model has an additive trend, and the Tableau model has a multiplicative trend. The last 20 years of CO2 data were deleted from the CSV file, which were provided to the various tools. The software predicts the CO2 data 20 years into the future, and the predicted values are then compared to the observed values. Tableau and PowerBI offer the option to export the forecast as a table. Cognos does not have this option, and the user must hover over each data point in the line chart to get the value of the forecast. For each year, an assessment is made based on the MAPE or the RMSE measurement. By evaluating the performance for each year, further conclusions can be derived. Figure 3 shows the performance based on the MAPE measurement over 20 years.

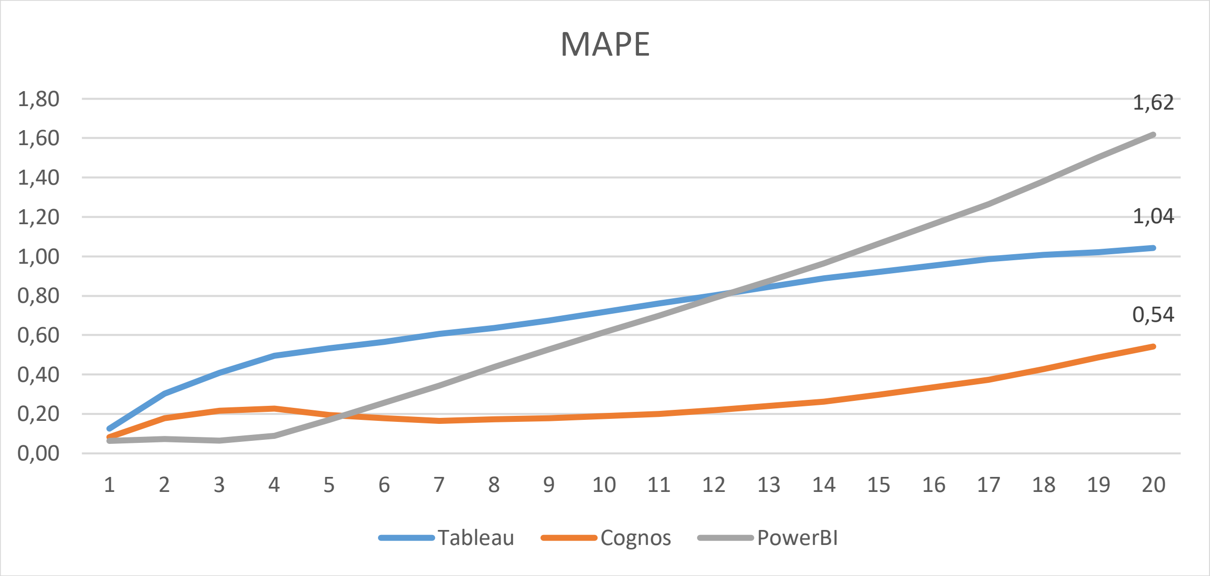


Figure 3: Forecasting evaluation based on the MAPE

In the first five years, PowerBI makes the best predictions. After the 5-year mark, Cognos has the lowest MAPE until the end of the 20-year forecasting span, at 0,54%. Around the 12-year mark, Tableau starts to perform better than PowerBI and ends up with a percentage error of 1,04%. PowerBI has a MAPE of 1,62% after the 20-year forecasting horizon. The next figure shows the performance based on the RMSE measurement.

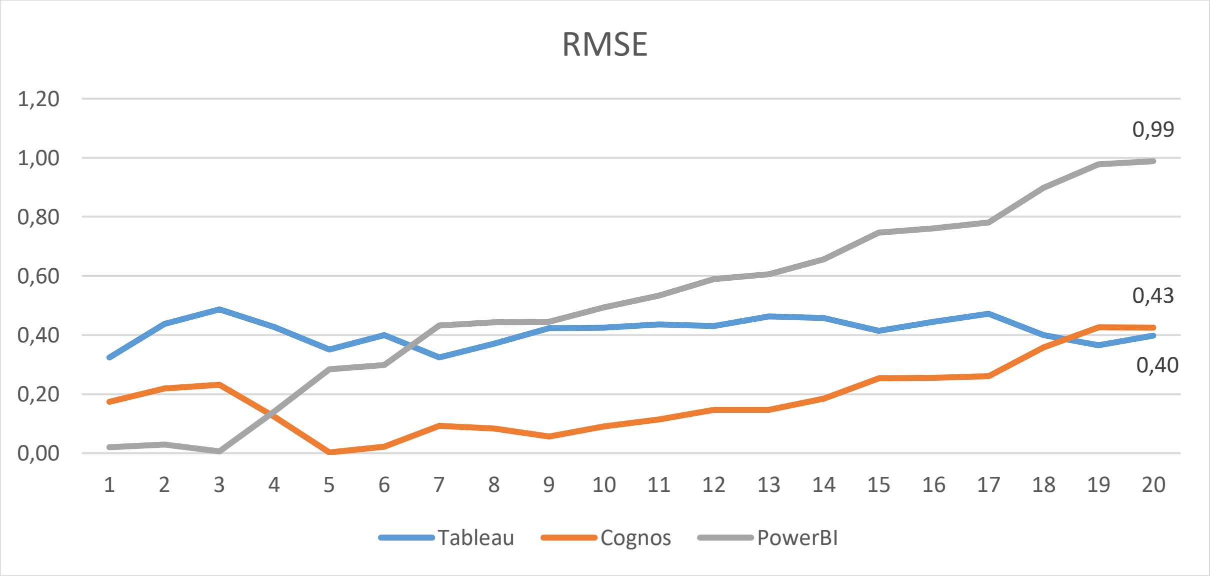


Figure 4: Forecasting Evaluation based on the RMSE

PowerBI also shows the best performance on the RMSE measure till year four and has a higher score than Tableau till around year seven. Cognos has a significant low RMSE of 0,002 in year 5. Cognos and Tableau converge by the 20th year of forecasting, with Cognos ending up at 0,43 and Tableau with 0,40. PowerBI has a RMSE of 0,99 at the 20-year mark.

Regarding the implementation of forecasts in each model, figure 5 displays the forecasting options in each of the software. All software offer the ability to change the forecast horizon, the time periods to be ignored for the forecast and the confidence interval. PowerBI and Cognos offer the ability to change the seasonal period, while Tableau can fill in missing values with zero and change the model to a custom model. The custom model can change the seasonality and trend to either "none," "additive", or "multiplicative”. The implementation of the prediction and the evaluation of the process will be done simultaneously in the next chapter.

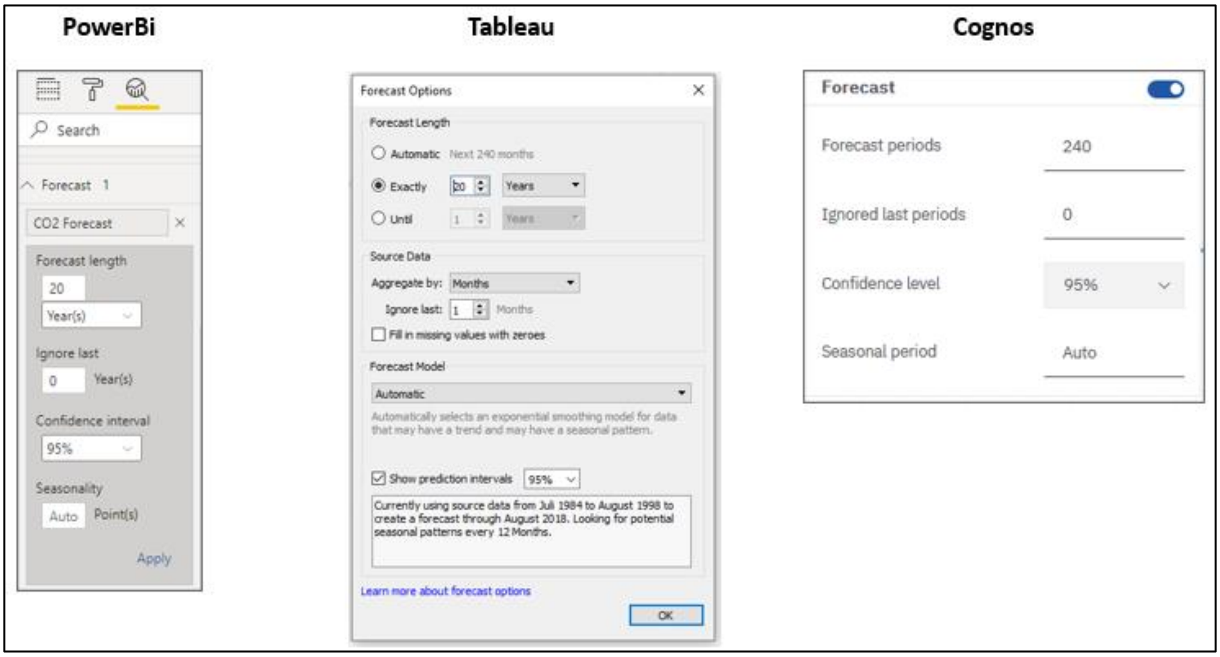


Figure 5: Comparison of forecasting options

5. Discussion

The discussion addresses aspects of the two research questions. First, the prediction performance of the three software programs is analyzed based on the previous findings. Second, the CO2 prediction implementation process is modelled and discussed.

5.1 Performance analysis

Tableau and Cognos provide the user with the ability to examine the forecasting model in detail. While the trend in Cognos is modelled additively, Tableau models the trend multiplicatively, i.e., the trend is multiplied by the other components. In addition, the alpha coefficient of the Cognos model is 0.033 higher than the alpha coefficient of the Tableau model, meaning that Cognos relies slightly more on recent data for level smoothing. The beta coefficient, responsible for trend smoothing, is 0.025 higher in the Tableau model. The largest difference in coefficients is the gamma coefficient, which indicates seasonal smoothing. Cognos' gamma coefficient is 0.234 and Tableau's is 0.108, more than half compared to the Cognos value. As a result, the Cognos model relies much more heavily on current data for seasonal smoothing compared to Tableau. The seasonal time frame for both models is 12 months, which means that the seasonal patterns are annual.

The predictive performance of the model, provided by the software, is similar. The MASE and MAE values are better for the Cognos model, the RMSE value is better for the Tableau model, and the MAPE value is the same for both models. These values give the user a sense of how good the forecasting model is. It is evaluated based on shorter forecasts and then compared to observed values. These details of the models are a good way for the user to inspect details of the forecast model. In addition, Tableau provides a link below the details for more information on the exponential smoothing model. It leads to Tableau's web page, which provides further descriptions of the forecasting methodology and the model. Cognos does not provide a direct link, but still has information about forecasts and models on its website. PowerBI not providing details or information about the model used is a missed opportunity and could disqualify them for scientific use. If predictions are used in a scientific setting, model details must be provided.

Performance evaluation based on MAPE and RMSE measurement over a 20-year period was shown in the two graphs in figure 3 and figure 4. For both measurements, PowerBI provided the best predictions in the first three to five years. After the initial time frame, PowerBI showed a steady upward trend in both measurements. This upward trend is steepest for both measurements compared to the other two software. This indicates that the upward trend continues as the forecast horizon increases, which means that PowerBI performs best compared to Cognos and Tableau for short-term forecasts, but worst for long-term forecasts. In both measurements, Tableau starts with the worst forecast performance. Notably, there is no upward trend in the RMSE measurement, i.e., performance remains at the same level over the 20-year period. For the MAPE measure, the upward trend also slows over time. With the trend in MAPE measurement, Tableau would eventually pass Cognos if the forecast horizon were increased further. These results suggest that Tableau has a low short-term forecasting performance but remains more static compared to Cognos and PowerBI.

The Cognos model starts second in both measurements but begins to perform better after about three years. In particular, the RMSE measurement shows a steep decline with a minimum in the fifth year. Cognos is the best performing software over time in both measurements. Since in the RMSE measurement the level of the observed and predicted values plays an important role and the MAPE is the percentage error, the MAPE can be considered as a more reliable measurement. Therefore, the Cognos model is not the best model in the first years, but it outperforms the other software programs over a 20-year period. The CO2 concentration in the atmosphere will increase in the future. Climate-related topics are often studied in the long term. Therefore, long-term performance is preferred when studying CO2 data with different software solutions. Cognos can produce the most reliable forecasts compared to Tableau and PowerBI. Looking at the final MAPE value over the 20 years, Cognos, Tableau and PowerBI (0.54 %, 1.04 %, 1.62 %) show high quality forecasts. These forecasts were made with exponential smoothing. However, the performance, especially in terms of the MAPE, shows that the models were implemented in different ways. None of the software provides an alternative model for forecasting. These alternatives could be ARIMA or machine learning techniques, e.g., decision trees or neural networks.

In summary, with a maximum absolute percent error of 1.62 % over a 20-year forecast horizon, all three forecasting tools performed well. With PowerBI performing particularly well in the short term, Cognos having the best overall performance, and Tableau having the lowest upward MAPE trend, all software programs have their advantages. This indicates that analytical software are able to forecast seasonal and trend data with a high accuracy. To what extend forecasts of different datasets have a high quality, is uncertain.

5.2 Forecasting process analysis

Appendix 1 depicts the process of each software to achieve the 20-year forecast of the CO2 data. The process can be divided into import, forecast, and export. The process of each software is discussed in terms of intuitive behavior.

Tableau has the most process steps of all software programs. The import process is not trivial. The software does not provide any onboarding help. By right-clicking on the CSV file after importing it into Tableau, the user can edit the properties of the text file. The file separator and location setting must be changed to alter the decimal separator options. The user needs to know which country uses the decimal separators used in the CSV file. A better approach would be to have a decimal separator option. Tableau is the only one of the three software to offer the option to display the forecasts not only as a chart, but also as a table. The forecast and forecast options can be set by right-clicking on the data, as shown in figure 6.

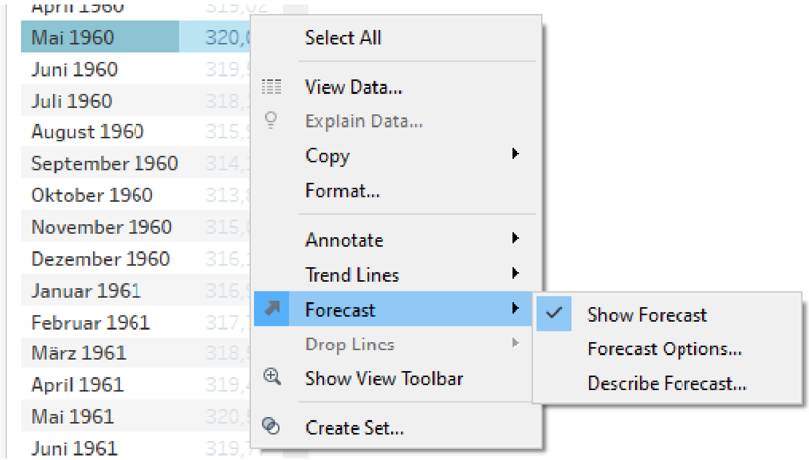


Figure 6: Forecasting options of Tableau

The forecast options are shown in the findings in figure 5. Tableau has a pop-up window for these options. The options are well organized, with a forecast summary at the bottom of the window. The forecast summary via the "Describe Forecast" option is also well laid out, with the option to copy to the clipboard. This option allows you to copy the details of the model into Excel or other tools.

Tableau offers the possibility to display and extract the data as a CSV file when the forecast is created in a table. This provides the user with the ability to use the data for further research. All in all, Tableau's import process can be time consuming depending on the amount of data due to the further research required. The forecasting process is structured, detailed, and user-friendly. Since Tableau is the only tool that allows forecasts to be run in a spreadsheet, it provides a useful additional future.

Cognos' import process is simple, as the user does not have to change any import options compared to Tableau or PowerBI. The software recognizes the data correctly. The forecasting process is also clear and straightforward. In Cognos, the model summary is only visible by dragging up the bottom portion of the screen. This option is barely visible and not indicated anywhere on the page. The biggest drawback of Cognos in terms of forecasting is the lack of ability to export the data. The only way to obtain the data is to hover over the data point and manually enter the forecast value into Excel, for example. This disqualifies the software for any forecasting efforts, as it is not practical to manually transfer the data each time.

PowerBI does not provide the option to change the location settings when importing the CSV file. Therefore, the regional settings must be changed prior to importing the file. This process of changing the location beforehand needs to be researched by the user. This process of importing the CSV file is not intuitive and is a disadvantage. PowerBI automatically converts the data so that it is totaled by year, not by month as indicated in the original data. This necessitates the "expand all down in the hierarchy" option, which is not clearly visible to the user at first. The export option is useful for further evaluation of the forecasted values.

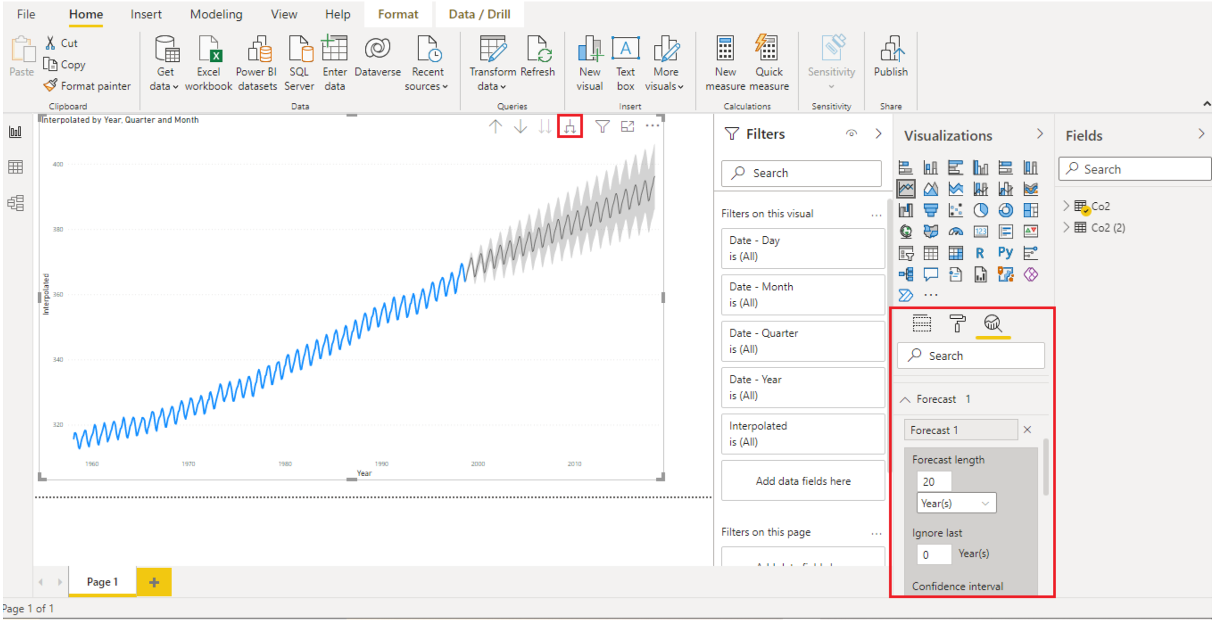


Figure 7: Forecasting option in PowerBI

In summary, Cognos offers an intuitive process of implementing forecasts, especially when importing the file, in addition to enhancing model details. However, the lack of the ability to export forecast data limits the software to visualizing forecasts. More advanced analysis of the data is not possible. PowerBI is overall solid for performing forecasts and exporting the data. However, options such as the option to change locations or the small forecast window are not user-friendly. The biggest drawback of PowerBI is the lack of model details. The user cannot see how the software calculated the predicted values. Tableau has the most implementation steps, but the most user-friendly interface, descriptions, and options. Especially the possibility to display the data in a table, model details and links for further information make Tableau the most intuitive software for implementing forecasts. Paired with a high forecasting performance makes Tableau the number one choice for forecasting, out of the three examined software. Tableau shows signs of performing better in the long-term perspective, which is particularly useful for climate data.

5.3 Implications, limitations and recommended future research

In this paper, the prediction performance and the prediction process were analyzed. It was shown that only Tableau can be used in scientific research, as Cognos does not have an export option and PowerBI does not provide model details. Cognos and PowerBI could be used as a visual element in practice, but not as a scientific forecasting tool. Tableau also offers only one model, so its use for scientific research has to be well considered.

The limitation of this work is the lack of comparisons to other forecasting models and visualization software. Statements could be made about the relationship between the three examined software and the exponential smoothing model. The extent to which other models have better forecasting performance is not known. In addition, Gartner's magic quadrant for business intelligence platforms shows a variety of unique software. Because the scope of the work is limited, only three software could be compared.

Further research could include comparing the exponential smoothing model with various forecasting models such as ARIMA, a decision tree model, or neural networks based on the CO2 dataset. By comparing and identifying the best model, this work could be put in relations to the best predictive performance.

6. Conclusion

All in all, all three analytical software, PowerBI; Tableau and Cognos, are able to forecast atmospheric carbon dioxide data. With the help of the theoretical foundations that have been made in the literature review a profound analysis could be made. Nevertheless, all three tools have their advantages and disadvantages, and it is therefore important to define beforehand the needs and goals of a carbon dioxide forecasting in order to get the best result. Whereas Tableau can be used as rather scientifical software, PowerBI focusses more on the actual visualization of the data. Due to the automated selection of either an additive or multiplicative model of trend and seasonality, different software have taken different models. For example, Cognos has calculated the value on basis of an additive trend whereas Tableau selected the multiplicative model for the carbon dioxide data. Since several KPIs were used for the evaluation method it turns out that different measures predict different values. Through the different KPIs the performance of the analytical tools gets evaluated differently. Moreover, it is important to involve also other forecasting and evaluation models like ARIMA or neural network in the future to rate the performance of each software compared to other benchmark models to indicate a tendency.

The second research question focused on the intuitive process of the implementation of the dataset. Through the ascertainment of each process step a profound analysis could be made. It turned out that the analytical software have different features to implement the data. The most intuitive and shortest process of implementation was able with Cognos whereas with PowerBI three more steps had to be done before the data got successfully implemented. Additionally, the model had also limitations. Cognos was not able to export the data automatically and PowerBI did not show the used model.

For further research other analytical tools should also be analyzed. Even though the intuitive process has been rated objective there is always a subjective definition of intuitive. Therefore, it is also thinkable to evaluate the intuition through quantitative research by having a larger number of experts implementing the data. Furthermore, the available sources on the tools are rare and rather subjective since they are mostly from the companies itself. Nevertheless, through analyzing the data and implementing them itself it was possible to scientifically evaluate the stated research questions.

References

Abdullah, L., & Pauzi, H. M. (2015). Methods in forecasting carbon dioxide emissions: A decade review. *Jurnal Tekonologi*(75-1).

Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*(68-8), pp. 1717-1731.

Box, G. (2013). Box and Jenkins: Time Series Analysis, Forecasting and Control. In P. Macmillan, *A Very British Affair. Palgrave Advanced Texts in Econometrics.* (pp. 161-215). London.

Brockwell, P., & Davis, R. (2016). *Introduction to time sereis and forecasting.* Springer.

Conroy, D. E., Kaye, M. P., & Schantz, L. H. (2008). Quantitative research methodology. In T. S. Horn, *Advances in sport psychology* (pp. 15-30).

Data Hub. (2018). *CO2 PPM - Trends in Atmospheric Carbon Dioxide*. Retrieved August 12, 2021, from https://www.datahub.io/core/co2-ppm

Gartner. (2021). *2021 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms*. Retrieved August 17, 2021, from Gartner: https://info.microsoft.com/ww-Landing-2021-Gartner-MQ-for-Analytics-and-Business-Intelligence-Power-BI.html?LCID=EN-US

Global Monitoring Labatory. (2021). *Trends in Atmospheric Carbon Dioxide*. Retrieved from Global Monitoring Laboratory - Earth System Research Laboratories: https://gml.noaa.gov/ccgg/trends/ff.html

Hyndman, R. J., & Athanasopolos, G. (2018). *Forecasting: principles and practice.*

IBM. (2014). *Trend Forecast Formula*. Retrieved August 25, 2021, from IBM: https://www.ibm.com/docs/en/cognos-analytics/10.2.2?topic=SSEP7J_10.2.2/com.ibm.swg.ba.cognos.pwr_ppweb.10.2.2.doc/c_trendforecastformula.html#TrendForecastFormula

Jain, C. L., & Malehorn, J. (2005). *Practical Guide to Business Forecasting.* Graceway.

Kolassa, S. (2014, November 20). *What is the distinction between short term and long term forecasting?* Retrieved August 17, 2021, from StackExchange: https://stats.stackexchange.com/q/124844

Leach, S., & Weick, M. (2017, July 31). Can People Judge the Veracity of Their Intuitions? *Social Psychological and Personality Science*(9), pp. 40-49.

Mentzer, J. T. (1988). Forecasting with adaptive extended exponential smoothing. *Journal of the Academy of Marketing Science*(16-3), pp. 62-70.

NIST / SEMATECH. (2012). *Engineer Statistics Handbook*. Retrieved August 25, 2021, from https://www.itl.nist.gov/div898/handbook/index.htm

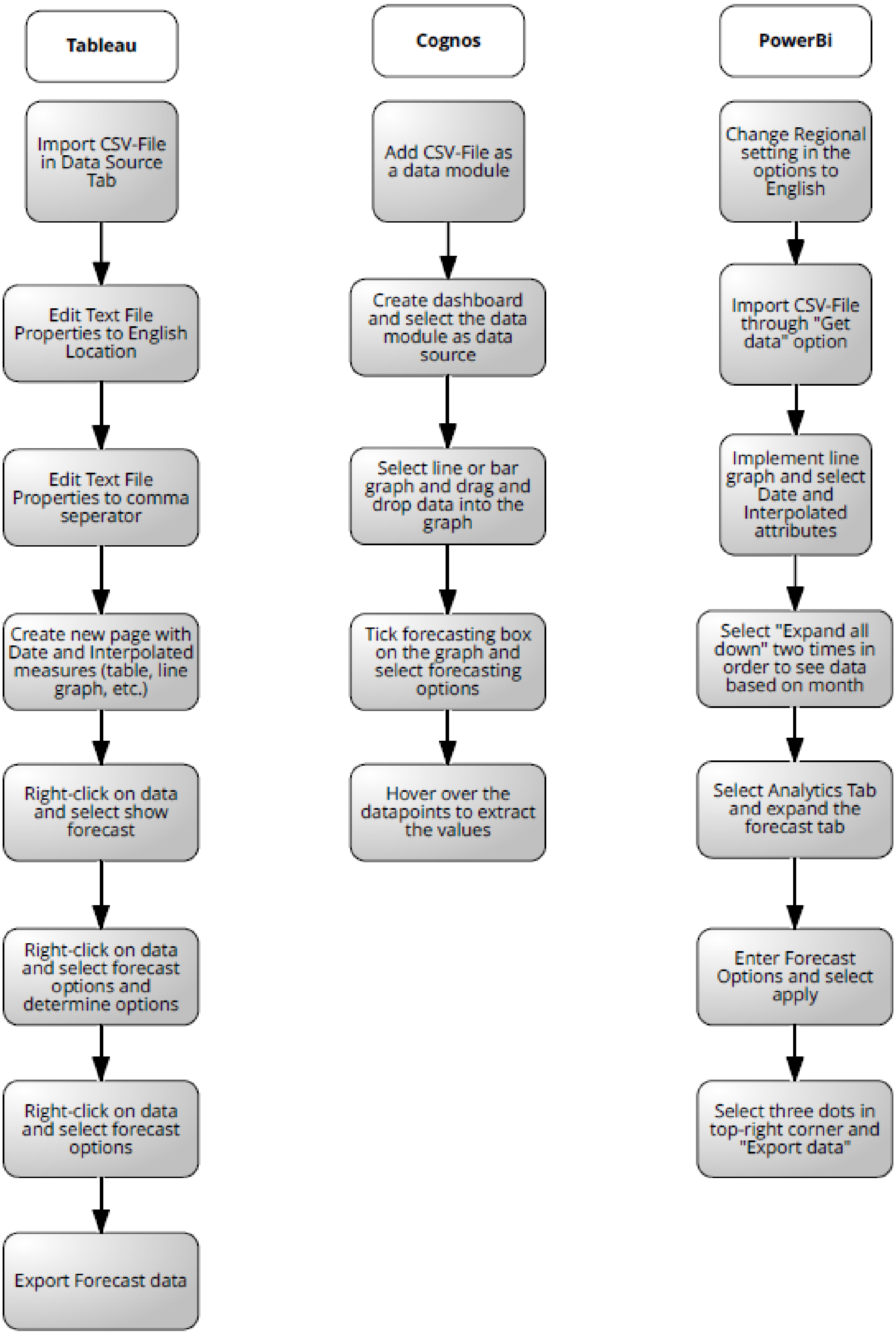
Richardson, J., Schlegel, K., Sallam, R., Kronz, A., & Sun, J. (2021). *Magic Quadrant for Analytics and Business Intelligence Plattforms.* Gartner.

Shcherbakov, M., Brebels, A., Shcherbakova, N., Tyukov, A., Janovsky, T., & Kamaev, V. (2013). A survey of forecast error measures. *World Applied Sciences Journal*(24-24), pp. 171-176.

Vandeput, N. (2019). *Forecast KPIs: RMSE, MAE, MAPE & Bias.* Towards Data Science.

Appendix

Appendix 1: Process of implementing forecasts



Appendix 2: Work distribution & number of words

Section	**Student**	**Words**
Abstract	Annika Neuß	268
Introduction	Prajwal Jagadish	948
Main part: literature review	Prajwal Jagadish	1,591
Main part: literature review	Annika Neuß	1,049
Main part: findings	Philipp Haid	1,399
Main part: discussion	Philipp Haid	1,803
Conclusion	Annika Neuß	388
Total		**7.446**

Declaration of Academic Intergrity

I confirm that this paper is solely my own work and that it has not been previously submitted for assessment as a whole or in part, nor published. All material which is quoted is accurately indicated as such, and I have acknowledged all sources employed fully and accurately. I agree with a plagiarism check of this thesis and know that the agreements of both expertises are necessary for a publication. Furthermore, I am completely aware that failure to comply with these requirements is a breach of rules and will result in resubmission, loss of marks, failure and/or disciplinary action.

Date: August 31, 2021

Signatures:
Annika Neuß
Prawjal Jagadish
Philipp Haid