BSc thesis in Industrial Engineering and Management

An Anomaly Forecasting Method for Support Processes

T. T. J. Leemans
2020

AN ANOMALY FORECASTING METHOD FOR SUPPORT PROCESSES

A thesis submitted to the University of Twente in partial fulfillment of the requirements for the degree of

Bachelor of Science in Industrial Engineering and Management

Commissioned by

CAPE Groep B.V.

University Supervisors

Lucas O. Meertens

Nils Knofius

External Supervisor

Bart Knol

By

T. T. J. Leemans

Jan 2020

Management summary

This method decreases obstruction of business operations by reducing unavailability of essential IT systems within companies. As the world, and especially businesses, get ever more dependent on computer services, unavailability can induce disastrous consequences for companies as well as customers. Full time availability is becoming an industry standard and can make or break a company's competitive advantage.

Availability of systems is diminished by anomalies. Anomalies are deviations from standard behavior in, for example, use of system memory or capacity. When it fluctuates to an amount the system cannot handle, the system shuts down. It becomes unavailable and companies can no longer provide their services. It is uncertain when these anomalies are going to happen, so many companies are currently not able to prevent them. This method equips companies, which have a support process to maintain their systems, with an anomaly forecasting solution that enables them to anticipate on anomalies before they happen so issues can be avoided.

The method starts by giving an overview of requirements and necessities which the company has to fulfill and determine. Based on these, an anomaly forecasting solution is chosen which best matches their situation and needs. An implementation plan is provided to integrate the solution and finally a validation setup is proposed to ensure it performs as desired. This validation contains a case study at CAPE Groep B.V., their data is used and based on their requirements an MVP is realized.

To sum up, this method provides a way for companies which maintain IT systems to preserve availability by anticipating on anomalies with the use of forecasting. Companies who perform such activities are highly recommended to implement this method as it can lead to preventing fatal system unavailability.

Acknowledgements

My research would not have been possible without the aid and support of my supervisors from the University of Twente, Lucas O. Meertens and Nils Knofius, and my external supervisor from CAPE Groep, Bart Knol.

I am profoundly grateful to Lucas O. Meertens for his untiring efforts to increase the theoretical relevance of this research. During our sessions he came up with a lot of idea's which helped to improve the quality of the research itself, as well as, the applicability. He made me widen the scope of this research to the point that all similar companies that aspire to apply machine learning or anomaly forecasting can benefit from it.

In addition, I would like to express my sincere gratitude towards Bart Knol for his inexhaustible willingness to make this research a success. Throughout many meetings together we discussed the ins and outs of all ideas I came up with and Bart provided me with useful insights and feedback to how these could further be utilized. His critical questions made me think of many possibilities and problems which I otherwise could have overseen.

I wish to acknowledge the help provided by Nils Knofius and would like to thank him for proofreading this thesis and providing me with a great amount of feedback. With his extensive knowledge about machine learning he enabled me to improve the theoretical model and validation substantially.

Furthermore, I would like to thank Gilang Charismadiptya for helping me analyze system compositions and implement the solution which could then be used for validation purposes. I would like to thank Jaro van der Beek for proofreading many chapters of this thesis, discussing problems I ran into, and in general being great company while performing this research as a whole.

I would like to thank CAPE Groep B.V. for providing an encouraging environment to perform this research in, as well as great colleagues, and for covering the financial aspects.

Finally, I would like to thank my family and friends for their motivational support which kept me ambitious to further expand the research and inspired me to keep going when problems arose.

Table of Contents

List of figures

Abbreviations

1.   Introduction

The aim of this thesis is to design, develop, and validate a method for implementing anomaly forecasting within support processes. This chapter provides the problem's theoretical background (1.1), the problem identification (1.2), the research approach (1.3) and the research design (1.4).

1.1.   Theoretical background

This paragraph clarifies terminology (1.1.1 and 1.1.2) used within this report, discusses the commissioning organization (1.1.3), and describes the theoretical relevance of the research as a whole (1.1.4).

1.1.1.   Anomaly forecasting

Throughout this research, the term forecasting is used. Forecasting is the process of predicting the future based on historical data (BusinessDictionary, 2019) (Investopedia, 2019). It attempts to cope with the uncertainty of the future by looking for trends in past and present data (BusinessDictionary, 2019). This research narrows this concept down to the prediction of upcoming deviations in standard behavior. These are called anomalies.

Forecasting in this research uses the concept of machine learning. Machine learning makes computer systems perform tasks without explicit instructions (Rouse, 2019). For this it uses algorithms and statistical models which are based on sample data (Rouse, 2019).

1.1.2.   Support processes

Within this research, the term support process is a reference to the pursuits or activities of a company, or a department within a company, which aim to maintain computer systems or applications. While this method could be applied to other processes as well, this report does not focus on or facilitate these.

1.1.3.   Commissioning organization

This research is commissioned by CAPE Groep, in short CAPE. CAPE is an IT consultancy company located in Enschede. Their main business is helping clients realize their business goals with the use of digital transformation (CAPE Groep B.V., 2019). They deliver tailor-made applications and integrations for companies which mostly operate in construction and logistics.

All work realized by CAPE is maintained by the support team. Their job is to ensure applications stay available and up to date. This research is performed on behalf of and guided by them.

1.1.4.   Theoretical relevance

Anomaly forecasting is of greatest importance for almost all companies in every type of industry (Heidemann, 2018). Prevention of downtime is a must since companies keep getting more dependent on their computer systems. However, the implementation of forecasting is not easy, which creates a need for methods that enable companies to do so. This method will help companies to reduce damages.

This project plan provides an approach on how to implement forecasting within a support process to improve application availability and uses the system composition at CAPE for validation purposes. However, the theoretical relevance of this research is wider than just CAPE. The research aims to provide a method for all companies like CAPE with a method on how forecasting can be applied within their own processes. As CAPE is the commissioning company, their composition is used to validate the method: within the method decisions are based on their requirements. Other companies can use it and base decisions on their own requirements.

1.2. Problem identification

This paragraph discusses the problem cause (1.2.1), the problem cluster (1.2.2), the core problem (1.2.3), the proposed problem resolution (1.2.4), the current versus the desired situation (1.2.5), and the research question (1.2.6).

1.2.1. Problem cause

Companies endeavor to provide the highest quality of services. Within a support process this means delivering 100% service availability for all customers. Based on interviews with employees at CAPE, the following causes for downtime were determined:

All systems or applications in the world contain flaws. This is because they are developed by humans and people make mistakes. These flaws make the applications induce incidents from time to time, decreasing the availability. Often, these flaws are fixed straight away but reaching a "perfect" application is still impossible. As developers keep innovating to further fulfill the customer's desires, new flaws will keep arising.

The availability of systems or applications is diminished by anomalies. Anomalies are deviations from standard behavior in, for example, use of server memory or capacity. When it fluctuates to an amount the server cannot provide, the server shuts down the application. The application becomes unavailable and a company can no longer provide their aimed quality. They do not know when these anomalies are going to happen so cannot prevent them.

Over the years, the number of systems or applications maintained by a company grows. This comes with an equivalent amount of workload for the support process. A company cannot hire a proportionate amount of personnel as this is not a scalable solution, while it is expected that in the future the amount will keep increasing with ever growing numbers.

1.2.2. Problem cluster

The in 1.2.1 stated problems plus their respective cause and effect relationships are visible in the problem cluster available in Figure 1.
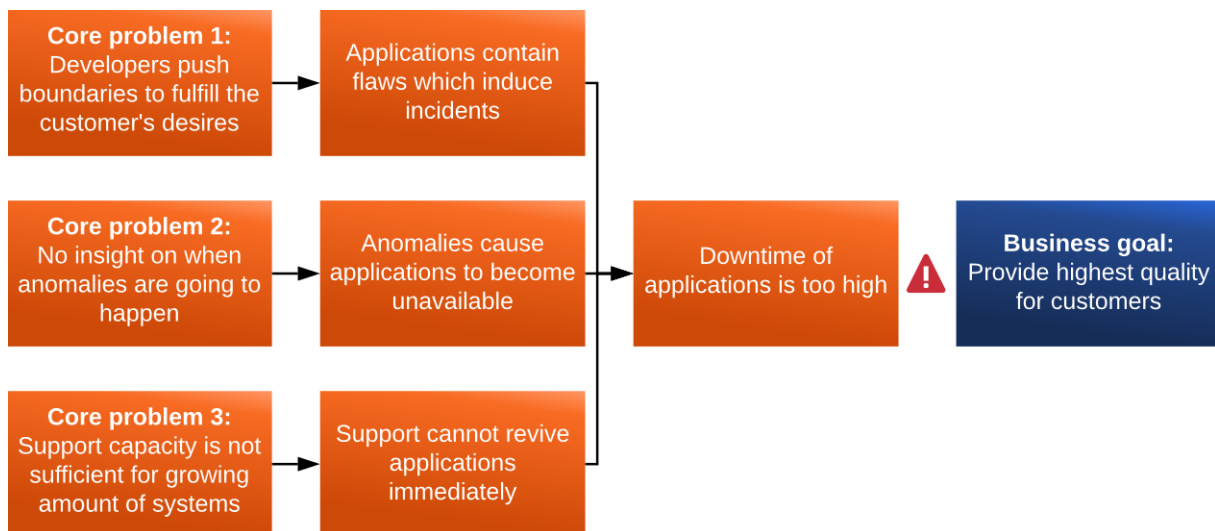

Figure 1: Problem cluster

1.2.3. Core problem

As visible in Figure 1, three core problems exist. Reasons for solving each of these are shortly discussed in this chapter. As this research is commissioned by CAPE, the most promising one

for them is chosen as the core problem for this research. It is decided based on their business goals and former investments.

Core problem 1: *"Developers push boundaries to fulfill the customer's desires."* comes with the business goal of CAPE to always strive for optimal customer satisfaction. It is one of their competitive strengths. To enable this, the CAPE academy constantly attempts to further develop the employees. Thus, CAPE is already heavily engaged in solving this problem.

Core problem 2: *"No insight on when anomalies are going to happen."* has at this time of writing not been investigated thoroughly yet, it also does not intervene with the business goals of the company.

Core problem 3: *"Support capacity is not sufficient for growing amount of systems",* comes with the aim of CAPE to keep growing as a company. The amount of applications which require maintenance will keep increasing therefore as well. On top of that, they even want to further invest in offering maintenance as a service. Since they want to keep this growth pace and provide their services at a competitive price, hiring more personnel is not a scalable solution.

Based on Heerkens (2017), when choosing which core problem to tackle one chooses that which has the greatest impact at the lowest cost. For this case, solving problems 1. and 2. will have the greatest impact as those can prevent downtime from happening, while 3. only means incidents can be recovered faster. Finding a solution for number 1. will incur way higher costs as the company is already heavily investing in solving this problem. For number 2. however, not much has been investigated yet on how this problem can be tackled. Thus, a lot of ground can be made with a relatively small investment. To conclude, at the time of writing, problem 2. is ought to be the leading core problem which need to be solved.

1.2.4. Proposed problem resolution

To solve the aimed core problem determined in 1.2.3, the following resolution is proposed: To increase the availability, measures which prevent the systems from becoming unavailable during anomalies must be implemented.

Such measures could, for example, comprehend rebooting systems or increasing capacity, but this depends on the metric for which the anomalies are being forecasted. To realize this, the support team must get an insight on when these anomalies are going to happen and what they will induce. This way, the appropriate actions can be determined beforehand and put into practice without the need for disproportional doomsday preparations, which cost money and resources. This insight can be achieved by means of automated anomaly forecasting. This research therefore proposes a method for implementing anomaly forecasting within the support process.

1.2.5. Current versus desired situation

To measure the success of this research the current and desired situation are stated. As this research aims to benefit both the commissioning company as well as the academic world, two current and desired situations are discussed:

Within the company, at this time of writing, no anomalies are forecasted yet. Anomalies can only be forecasted based on the intuition of employees, but these are based on known deviating time periods like holidays. With respect to other (unknown) patterns, the current situation is that 0% is being forecasted. The desired situation depends on the metric for which the forecasting is required, a metric and corresponding desired situation are discussed in paragraph 2.5.

At this time of writing, already a lot of investigation has been done for anomaly forecasting in application maintenance. Meanwhile, no method which is purely focused on implementing anomaly forecasting in support processes within companies which are similar to the one at CAPE Group, and that makes use of machine learning exists yet. At this point they are most often limited to the conventional forecasting methods such as moving average and exponential smoothing. Therefore, for this problem the current situation within the academic world is equal to the results which are achievable by using such statistical methods. The desired situation is stated as more than this current situation while this implies an improvement has been achieved.

A validation will make the results of this research measurable to see if the goal has been reached. The validation method is discussed in paragraph 2.5.

1.2.6. Research question
Based on the core problem determined in 1.2.3, the research question is formulated as follows:

How can anomalies be forecasted within a support process to increase application availability?

1.3. Research approach
This paragraph discusses the used methodology (1.3.1) for performing the research as well as the deliverables (1.3.2).

1.3.1. Methodological framework
To execute the research, the Design Science Research Methodology is used, in short DSRM. This methodology is used while it focusses on realizing an artifact (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). This is necessary, as an MVP is needed to validate the findings. The process model belonging to the DSRM contains 6 phases, these are visualized in Figure 2.
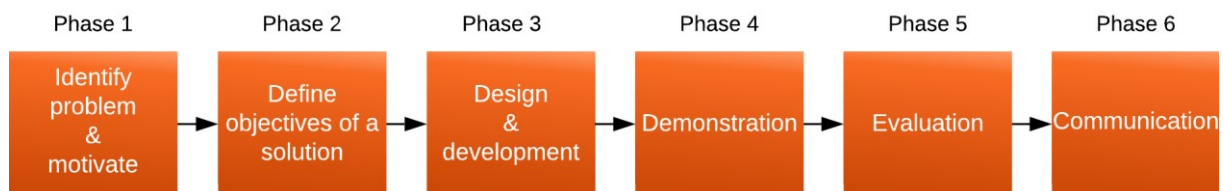
Figure 2: DSRM process model (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007)

In phase 1 the problem is defined as well as the value of a solution motivated. Phase 6 comprehends the report and the presentation of the results. Within phases 2 to 5, sub questions must be answered. These are discussed next.

Phase 2, define objectives of a solution
At the time of writing, it is unclear what is required from a forecasting system. This must be determined to find a suitable solution. Thereafter, the default solutions which are already available on the market must be found and compared based on these requirements:

1. *Which anomaly forecasting solution fits the needs of a company best?*
   a. *What is required from the forecasting solution?*
   b. *What solutions for anomaly forecasting are available?*
   c. *Which of these solutions fits what company requirements or composition best?*
2. *What metrics are worth forecasting?*
3. *How can the performance of a forecasting setup be validated?*

Phase 3, design & development

It is unclear how the found solution must be implemented within the support process:

4.   *How can an anomaly forecasting solution be implemented in a support process?*
    a.   *How does a (default) support process look like?*
    b.   *How can a forecasting solution be implemented within such a process?*
    c.   *How can the results be communicated to the company?*
    d.   *How can be determined if the solution's results are valid?*

Phase 4, demonstration

The performance of the method must be validated:

5.   *How can the anomaly forecasting solution be implemented at CAPE?*
    a.   *How does CAPE's current system composition look like?*
    b.   *What system composition is applicable to CAPE's data?*
    c.   *How can the corresponding solution be implemented at CAPE?*

Phase 5, demonstration

The performance of the method must be validated:

6.   *How does the method perform?*
    a.   *Are the forecasts accurate when compared to the company's norm?*
    b.   *How does the accuracy of the forecasts compare to those of conventional statistical methods?*

Phase 6, evaluation

Results and advice for the company must be given:

7.   *How can the performance be evaluated?*
    a.   *What conclusions can be drawn?*
    b.   *What recommendations can be given?*
    c.   *What future research is required?*

For the phases for which they apply, the chapter within the thesis, the data gathering method, and the corresponding deliverable are available in Table 1.

DSRM Phase	Thesis chapter	Data method	Deliverables
1. Identify problem & motivate	Problem identification	Unstructured interviews, literature review	Problem definition, value of solution, problem approach
2. Define objectives of a solution	Theoretical framework	Literature review	Conceptual framework of relevant literature
3. Design & development	Design & development	Literature review	Implementation plan
4. Demonstration	Demonstration	Implementation plan, Unstructured interview	MVP
5. Evaluation	Evaluation	Validation setup	Validation of the MVP
6. Communication	Conclusion & recommendations	-	Conclusion and summary of research

Table 1: Phases with corresponding chapters, data methods, and deliverables

1.3.2. Deliverables

This research aims to realize the following deliverables:

- Report with advice for companies on how to implement forecasting within their support process. Containing:
    - Problem definition, value of solution, and problem approach
    - Summary of relevant literature
    - Implementation plan
    - Validation of the MVP
    - Conclusions and recommendations
- Minimum viable product which demonstrates and validates the functionality of the advice.

1.4. Research design

This paragraph contains the design of the research. As this plan is made for design science, most of the research design is already defined in the research approach. This chapter however discusses shortly the research method (1.4.1), the operationalization (1.4.2), data collection methods (1.4.3), and limitations (1.4.4).

1.4.1. Research method

This report contains a mostly descriptive research. It focusses on describing what solutions are available for anomaly forecasting, what the current situation within a default support process is, and how a solution can be implemented in such a process. The goal is to use the amassed and described knowledge for practical purposes.

1.4.2. Operationalization

The key variables within this research are related to the performance of the method, as well as, the validation of the, to be developed, MVP. As further discussed in paragraph 2.5, the validity of the method is partly dependent on the used literature and partly on that of the MVP. For the MVP this is an error evaluation measure, which value gives an indication of the validity of the MVP. These measures are discussed in paragraph 2.5 as well.

1.4.3. Data collection methods

To gather data, three different kinds of methods are used. These were already visible in Table 1. The first are unstructured interviews with employees at CAPE. This is required to gather intel on their requirements, and on how a chosen solution can be implemented. A second is literature review. Here, articles are analyzed to get intel on what solutions are already available on the market. The third is by means of a validation setup. The MVP this research aims to realize will be used to validate the proposed method. It is validated based on the key variables described in 1.4.2.

1.4.4. Limitations

The research is limited to the following:

- The research is executed at CAPE Groep B.V. Enschede, their requirements and process are used to determine the validation of the final method. The research is however usable for all kinds of companies who aim to forecast anomalies and have similar datasets.
- The requirements set for the datasets are based on those at CAPE at the time of writing. These can however change over time and are therefore not fixed. The described situation is only used for validation purposes.
- The data used for the validation of this research is not related to any of CAPE's current customers to preserve confidentiality.

2.   Theoretical Framework

This chapter discusses the requirements which a solution must comprehend (2.1), the available solutions with respect to forecasting techniques (2.3), what solution fits what system composition best (2.3), and the techniques used for the validation (2.5).

2.1.   Requirements

To limit the available forecasting solution to those which are applicable to this research a list of requirements that the solutions must comprehend is provided. Each of these helps to ensure successful anomaly forecasting:

1. Able to detect and forecast anomalies: *Basis of the solution.*
2. Ability to read and access historical data of company: *Needed to train solution for accurate forecasts and detections.*
3. Able to save data to a data source: *Make forecasts permanently available.*
4. Able to communicate forecasts to user: *To make user aware of anomalies. For example, via a dashboard.*

2.2.   Existing solutions

For anomaly forecasting multiple procedures exist (Galvas, 2016). To fully understand the concept of forecasting and to choose the most suitable approach, a description is provided as well as a derivation of the possible approaches and corresponding available solutions.

2.2.1. Forecasting techniques and approach

Forecasting itself is a collective term for all activities of making prediction about events in the future (BusinessDictionary, 2019) (Investopedia, 2019). It can be applied within many subjects, for example weather or stock prices, and done by means of different techniques. As stated in paragraph 1.2, this method focusses on forecasting anomalies. To do so, it makes use of artificial intelligence which is further elaborated in paragraph 2.2.2. A schematic overview is provided in Figure 3.
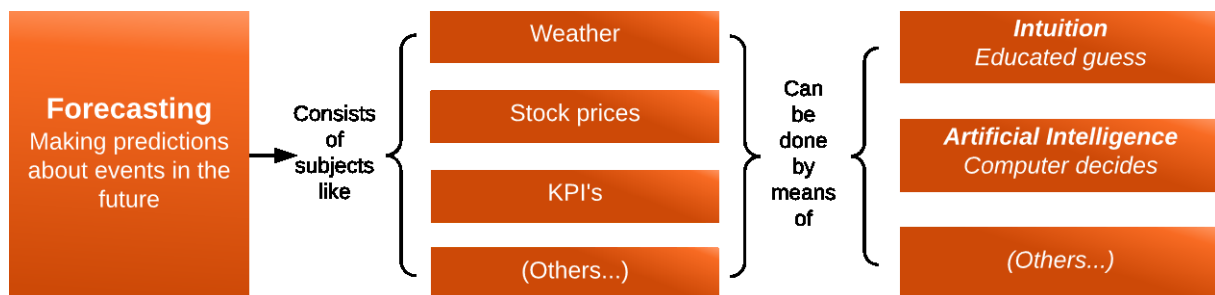
Figure 3: Theoretical model with respect to forecasting subject and method

2.2.2. Artificial intelligence

Artificial intelligence, hereafter AI, is an umbrella term for enabling computers to make decisions based on sets of rules (Touger, 2018). For this method, AI is used to carry out the forecasts. This is chosen because there can be assumed that using computers is a cheaper method than to make humans do it, as in general, buying more computer capacity is cheaper than hiring more personnel (Chowdhury & Sadek, 2012). On top of that, AI is more scalable, if the number of indicators for which forecasts are requires grows, hiring a proportionate amount of personnel can become undoable while a computer can be expanded with ease (Chowdhury & Sadek, 2012). Hence, making a computer do the work is chosen as the favorable solution. A schematic overview is provided in Figure 4.
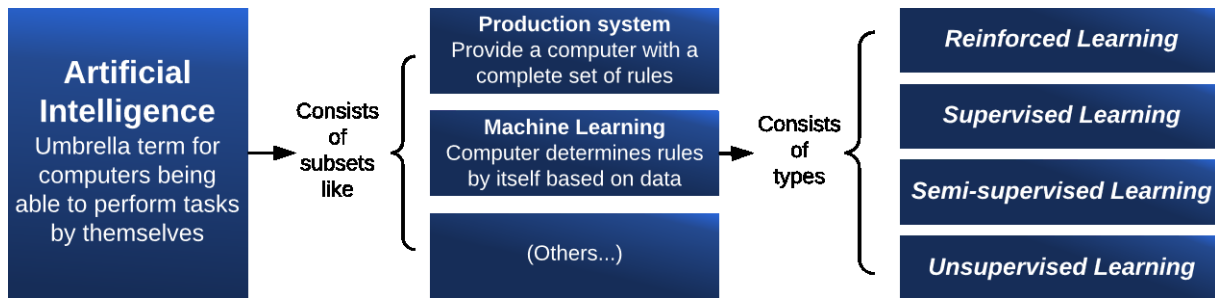
Figure 4: Theoretical model with respect to Artificial Intelligence and Machine Learning

2.2.3. Machine learning

Within AI many subsets exist, among them is machine learning (Nicholson, 2019). Machine learning, in short ML, is the field of science in which a computer is trained to make decisions based on data (Nicholson, 2019). It determines rules by itself that otherwise would have to be established by humans. There can be assumed that the need to set rules for every forecasting model will become undoable when the number of required forecasts increases, hence machine learning is preferred over a production system.

As visualized in Figure 5, the research area of machine learning is not only a subset of AI and can be used for forecasting, but also has overlap with statistics. Yet, the decision has to be made whether to use conventional forecasting methods using statistics or machine learning. There is no clear distinguish between the two, as it is a matter of perspective and differs based on the person you are talking to. The field of machine learning is however built on a statistical framework, since it uses data, and data has to be described using a statistical framework (Stewart, 2019). But whereas in statistics one must choose the underlying model of a dataset to perform the predictions, in machine learning the closest possible match to the behavior of the data is provided (Teboul, 2018). This will never be a perfect match and might not outperform these models, but it does take away the need to fit a model to each of the datasets (Teboul, 2018). As discussed, the number of forecasts is thought to increase significantly over the years thus machine learning is proposed to be a favorable option.
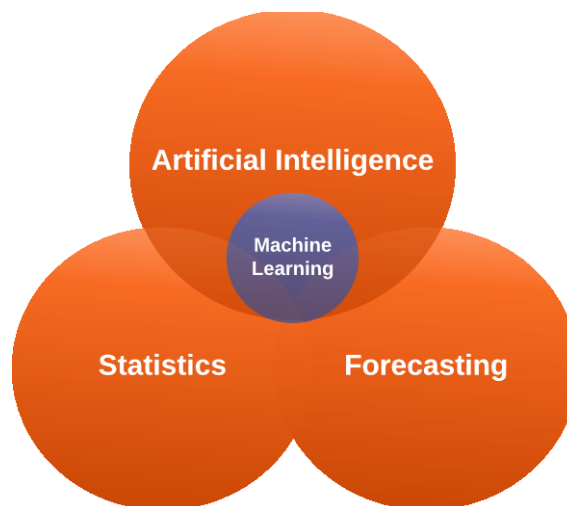
Figure 5: Correspondence of research areas

For a machine learning model to make accurate decisions, it needs a machine learning algorithm to train itself (Rashidi, Tran, Betts, Howell, & Green, 2019). Many different algorithms have been developed which can be classified into 4 different machine learning types (Grotov, 2019). These types are reinforcement learning, supervised learning, semi-supervised learning, and unsupervised learning (Grotov, 2019). To realize anomaly forecasting by means of

8

machine learning, one of these types must be chosen. The types differ in their requirements with respect to available training data (Grotov, 2019). As visualized in Figure 6, reinforcement learning has the highest requirements and has therefore the least amount of applicable data available (Grotov, 2019). Unsupervised learning, on the other hand, has the lowest requirements and has therefore the highest amount of available training data (Grotov, 2019). One should choose the type which fits the available training data best.
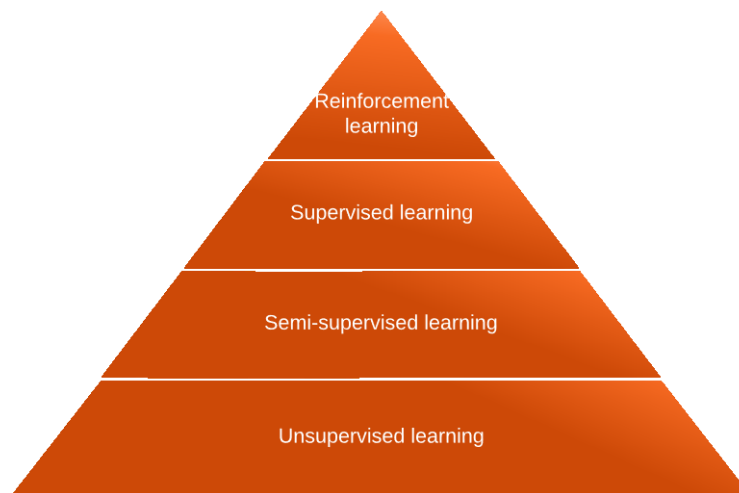
Figure 6: Availability of training data per machine learning type (Grotov, 2019)

Reinforcement learning requires some form of interaction environment to work, which is often a competitive opponent, like in a game of chess. The system trains by making decisions and evaluating the consequences. (Rashidi, Tran, Betts, Howell, & Green, 2019) Such an interaction environment must be developed by the company. Whether the effort to develop this environment is justifiable is assumed to depend mostly on the complexity of the environment and the necessity of using reinforcement learning. Whenever this environment is not available but other learning types give sufficient results, reinforcement learning is assumed to be an inferior option. Only when the results of the other learning types are not tolerable it becomes legitimate to develop one. An example of when this happens is when no historical data is available. Since, in contrast to the other learning types, reinforcement learning does not require historical data to work (Case, 2019).

Supervised learning is used when labeled data is available, this means, for each datapoint the corresponding decision made must be defined (Rashidi, Tran, Betts, Howell, & Green, 2019). For example: for each point in timeseries data, a column must state if the data describes an anomaly or not. When no labels are available, there can also be considered to label the data by hand. For example, by adding a label which states if an event is going to happen in a specific number of minutes from the current datapoint. Again, whether the effort of labeling data is justifiable is debatable and depends on the complexity and necessity. If the semi-supervised or unsupervised learning methods do not provide tolerable results one will have to start labeling data. The complexity of labeling data is dependent on the number of data sets. If the number of data sets grows significantly, one can assume it becomes undoable if all of these require a different labeling approach.

Semi-supervised learning combines supervised and unsupervised learning, it comprehends a middle ground of the two. For semi-supervised learning, some of the data must be labeled which is used to label the remaining unlabeled data to create pseudo labeled data (Zhu, 2005). Both are then combined to create a dataset which is applicable for supervised learning.

Unsupervised learning is the remaining type. This type is always applicable as long as enough historical data is available. It does not require data to be labeled and is therefore the remaining option when the requirements of the other options cannot be met. (Rashidi, Tran, Betts, Howell, & Green, 2019)

2.3. Matching solutions

Now it is known what is required of the solution, and what types of solutions are available. It is time to determine what solution matches a specific type of company. Firstly, the appropriate machine learning type must be determined based on the data which the support team has available to train the model as discussed in 2.2.3. For this, the flowchart in Figure 7 offers a schematic overview.
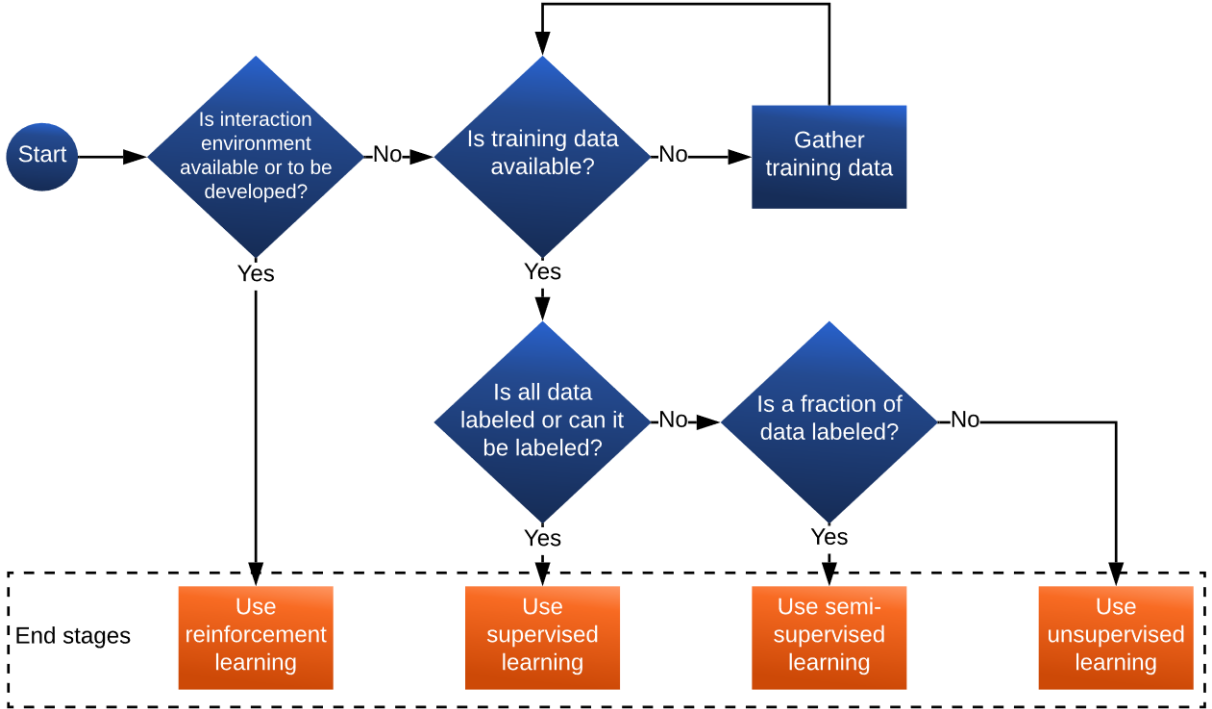
Figure 7: Flowchart for choosing machine learning type

The solutions are now limited to those which are applicable to the available data. To choose the solution within this set of machine learning solutions, the following starting point is proposed: Search for a solution which has stated compatibility with the data-source where the training data is stored. This due to the following: The foremost requirement for accurate forecasting is the accessibility of proper training data (Ray, 2015). If the developers of a solution have validated their solution, there can be assumed that it is usable as long as the data is fed towards the solution. Thus, when there is compatibility between the solution and the data-source, no alignment or further development is required. If no compatible solution is available, the training data is to be migrated to a data-source solution which does have such a compatibility. This is assumed to be easier and more likely to succeed than trying to adjust the solution. Figure 8 provides a schematic overview of this decision process.
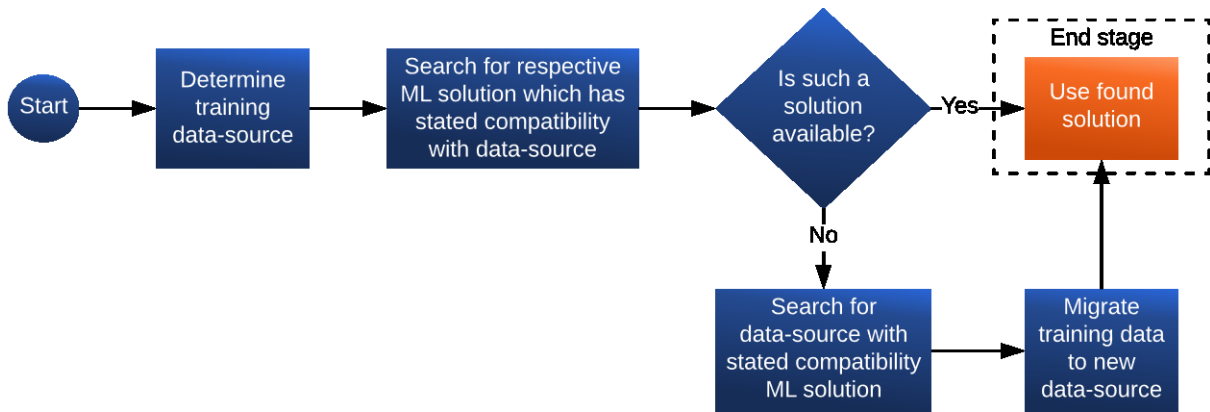
Figure 8: Flowchart for choosing solution

Now a suitable solution is found, the solution is ready to be implemented within the process. This is elaborated in chapter 3.

2.4. Choosing metric

When the forecasting setup is realized, it is time to choose metrics which are to be forecasted. This must be a well-considered choice as more metrics means a need for more server capacity. Too many metrics can also cause a user to be overwhelmed with data making them miss important results (Tanner, 2017). Forecasting too little metrics is also not an option as it might result in not using the full potential of the forecasting setup when important metrics are accidentally left out (Tanner, 2017). Therefore, only specifically chosen metrics are to be forecasted, and how to determine these is discussed next.

To determine the metrics that need to be forecasted the Goal-Quality-Metric (GQM) approach is used. This is a model where measures are approached from three distinct levels as represented in Figure 9 (Basili, Caldiera, & Rombach, 1994).

Figure 9: GQM approach

For a metric to be of need, it has to help reach a certain goal (Basili, Caldiera, & Rombach, 1994). Therefore, the first step is setting a goal which profits the company. Hereafter, a set of questions whose answers access whether the goal is being achieved or not (Basili, Caldiera, & Rombach, 1994). To answer these questions in a measurable way, respective metrics are chosen (Basili, Caldiera, & Rombach, 1994). These are the metrics which are worth being forecasted. An example is given in Figure 10.

Goal	**Quality**	**Metric**

Keep the CAPE information system accessible for employees → Can the system provide the required capacity? → Amount of free memory / Amount of CPU load / Amount of free storage space
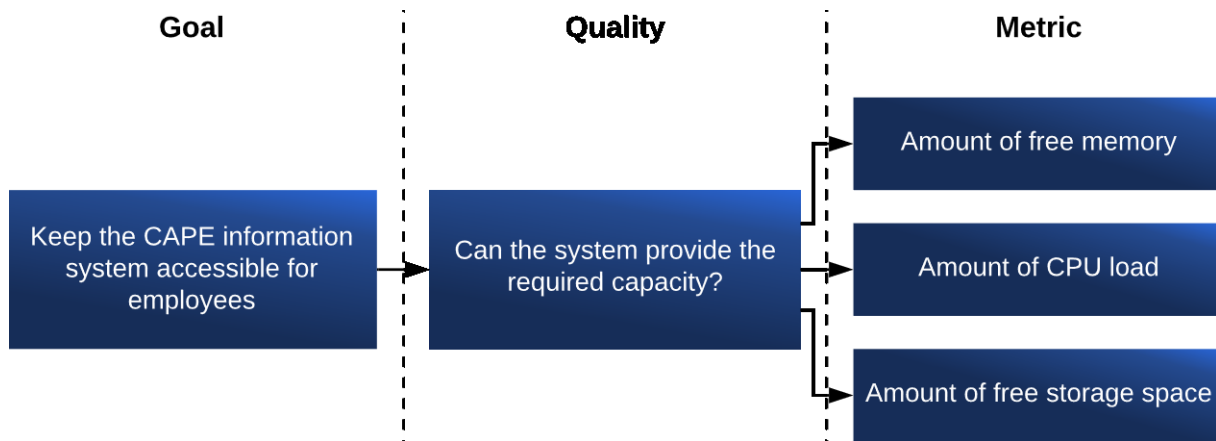
Figure 10: GQM example

2.5. Method validation

To ensure the legitimacy of the method, it must be validated. To do so, a similar approach to the one of the DSRM is used (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). This is chosen as this method is designed and developed according to the phases of the DSRM as discussed in paragraph 1.3.1. Peffers et al. (2007) state their method is valid based on three evaluation objectives: It should be consistent with prior research, it should be effective for the intended purpose of one or more cases, and it should provide a mental model for the presentation of research outcomes (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). This mental model is used for researchers to report the design and results of their research. While this third objective is not applicable for this method since this method is not developed to perform research, the others can. In the end of this paragraph, it is discussed how forecasts for the same period are performed with conventional statistical methods, and how these forecasts are compared to the ones from the machine learning solution.

2.5.1. Method validity

The first objective is to base the method purely on previous research (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). Their validity implies the validity of the resulting method. As this research is based on the DSRM itself as well as on literature which has been determined to be valid by the corresponding researches, the validity of this method can thus be deduced from those.

2.5.2. Case validations

The second object is to determine if the method is effective for the intended purpose of a case (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). For this method this means to determine if the forecasts are valid. To do so, an MVP is designed and developed. The validity of this MVP is determined as follows: The values for a random metric are chosen. For example, the one visualized in Figure 11.
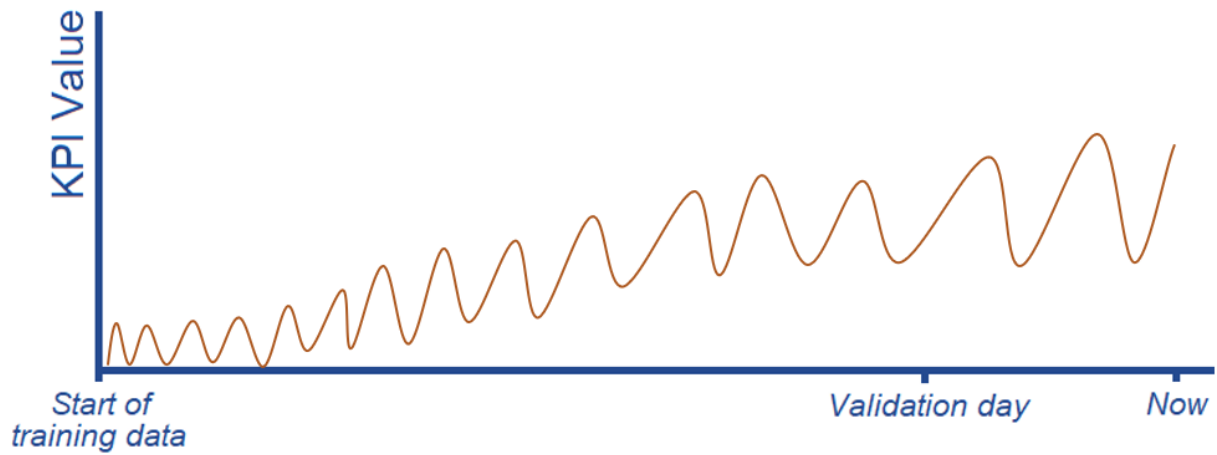
Figure 11: Plot of random metric

For the chosen metric, the data is copied and back until a randomly chosen "validation date" deleted. This time period is used as the forecasting period. Visible in Figure 12.
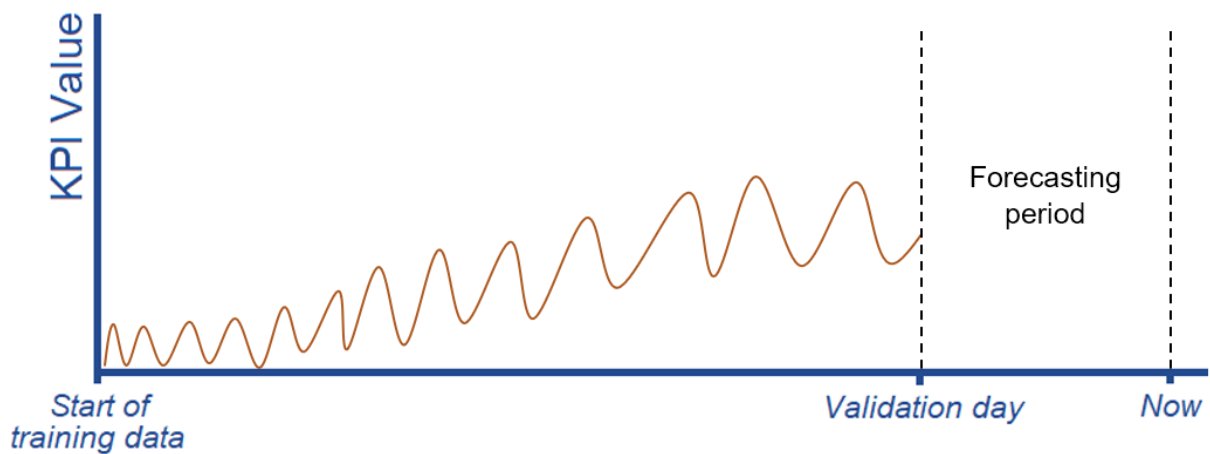
Figure 12: Same metric but with forecasting period removed

The MVP is trained with the remaining data and perform a forecast for the determined period. This forecast is then compared with the actual data which was deleted. This is visible in Figure 13.
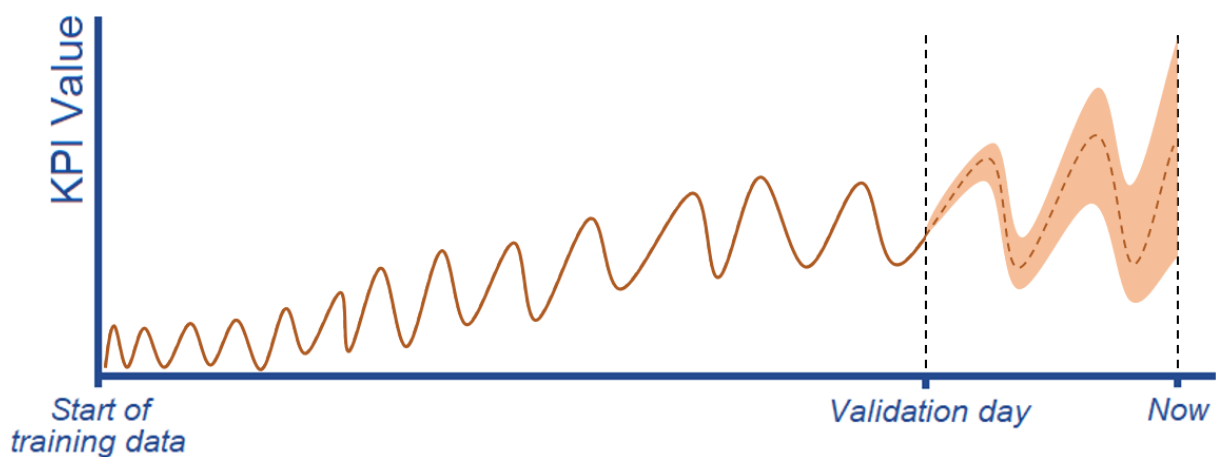
Figure 13: Random metric with forecast to be compared for accuracy validation

The forecasted data and the real data can now be compared. This is determined by two factors, the precision and bias (Vandeput, 2019):

- The bias is a representation of the historical average error and gives the overall direction of this error (Vandeput, 2019). For the forecasts this indicates whether they are on average too high or too low (Vandeput, 2019).
- The precision is a measure for the spread between the forecast and the real values (Vandeput, 2019). It gives an indication of the magnitude but not the direction (Vandeput, 2019).

As visualized in Figure 14, we aim for a precise and unbiased forecast.
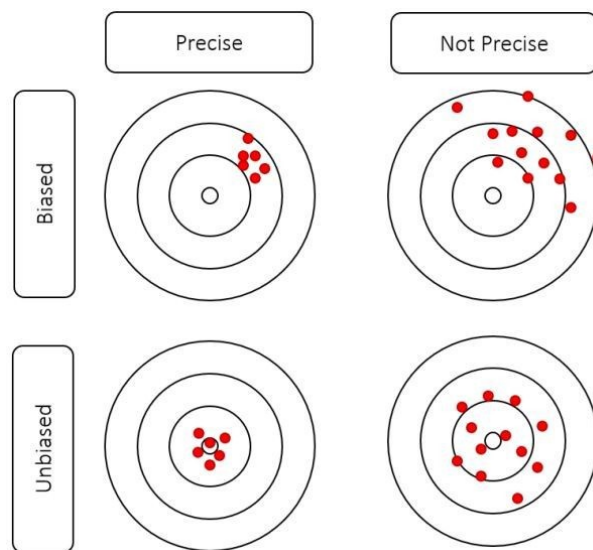
Figure 14: Precision and Bias (Vandeput, 2019)

The bias is calculated by taking the difference between the actual value and the forecast for each data point (Vandeput, 2019). This is called the error. The bias is the average over all errors. Vandeput (2019) states it can be calculated via the following equation:

$$Bias = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)$$

Here is A_t the actual (or real) value and F_t the forecasted value both at point t in time. The difference is taken which represents the error. When this bias turns out to be a positive value, the forecast is on average too low compared to the reality. And the other way around for a negative bias. (Vandeput, 2019)

For measuring the precision there is no one-size-fits-all indicator (Vandeput, 2019). Therefore, multiple evaluation measures are available (Lehmann & Casella, 1998), these are measures for the quality of estimators (Lehmann & Casella, 1998). Hence, in the case of this research, it can give an indication on how similar the forecast is to reality. The often used measures MAPE, MAE, RMSE, and MSE are therefore considered (Vandeput, 2019). How to calculate these is discussed next as well as when to use which. As no standard rule or value is available regarding these measures (Lehmann & Casella, 1998), there is chosen to compare them to a user defined maximum. This maximum is chosen based on what metric is forecasted. The user can then determine if the forecast is valid with respect to the user defined standard.

MAPE: As in businesses people often prefer to talk in percentages (Stellwagen, 2019). One can choose to calculate the Mean Absolute Percentage Error, in short MAPE (Stellwagen, 2019). Sellwagen (2019) states it is calculated via the following equation:

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

Here the error is calculated in the same was as when calculating the bias. This error is divided by the actual value to determine the proportion. The absolute is taken to remove any negative signs, and the mean is taken over all these proportions. Lastly, this mean is multiplied with 100% to make it a percentage. The dataset of actual values cannot contain any 0's as this will make the MAPE go to infinity due to the division by zero (Stellwagen, 2019).

MAE: When there is no preference for a percentage. One can choose for one of the other three options. The first to be discussed is the MAE, which is short for Mean Absolute Error (Swalin, 2018). Swalin (2018) states it is calculated via the following equation:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|$$

Here again the error is calculated the same as in MAPE but the error itself is used instead of the proportion. The absolute is taken to remove any negative signs, and the mean is taken over all these errors.

MSE: When there are no outliers (or it is unknown if there are outliers) within the dataset, the MSE can be considered instead of the MAE (Swalin, 2018). Short for Mean Squared Error, the MSE is calculated in a way that it emphasizes outliers (Swalin, 2018). Drakos (2018) states it is calculated via the following equation:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2$$

Again, the error is calculated the same as with MAPE and MAE. This error is squared to remove any negative signs and to emphasize the effect of the outliers (Swalin, 2018). It takes the mean out of all these squared errors.

RMSE: When the user needs the error to have the same unit as the metric for which it is calculated, the Root Mean Square Error is used (Drakos, 2018). MSE is used more often than RMSE as it is more easily to work with (Drakos, 2018). Swalin (2018) states is calculated the same as the MSE but now the root is taken as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2}$$

To provide an overview of when to use which evaluation measure, it is visualized in Figure 15.
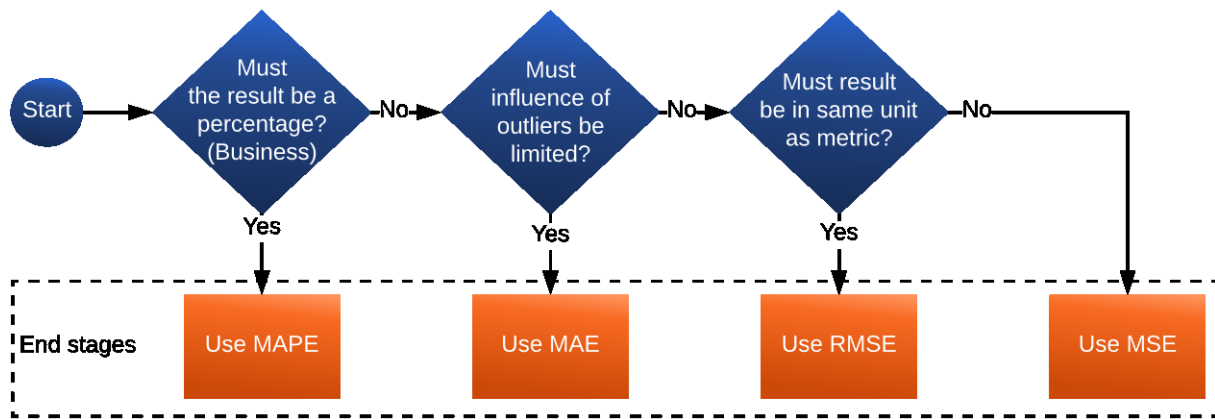
Figure 15: Flowchart for choosing evaluation measure

2.5.3. Conventional statistical methods

Now it is clear how the forecasts are evaluated; they can be compared to conventional statistical methods. This is used to evaluate how they perform against already often applied techniques and gives an insight on whether machine learning is the better way to go for the stated purpose of this research. It directly gives an indication whether an improvement has been achieved as discussed in 1.2.5. As stated by Nau (2019) nonseasonal patterns and trends can be extrapolated using a moving average or smoothing model. Hence, the chosen statistical methods are moving-average and exponential smoothing.

Forecasting by means of moving average is done for the value of a metric by taking the average over a chosen amount of the most recent observations (Nau, 2019). Nau (2019) states the moving average over the most recent m observations is calculated as follows:

$$F_{t+1} = \frac{A_t + A_{t-1} + \cdots + A_{t-m+1}}{m}$$

Here F_{t+1} represents the to be forecasted value at the next timestamp. A represents the actual value at the respective observation (Nau, 2019). The forecast is determined via Microsoft Excel and Excel's default interval of 3 recent observations is used.

As stated by Nau (2019) the forecasts of exponential smoothing are determined by taking the interpolation between the previous forecast and the most recent observation, where α controls the closeness of the interpolated value to the most recent observation:

$$F_{t+1} = \alpha A_t + (1 - \alpha)F_t$$

Again, the forecasts are determined via Microsoft Excel and Excel's default damping value for α is used which is 0.3. When the forecasts of both methods are determined, the evaluation measures as discussed in 2.5.2 are calculated for these forecasts as well. All for the same periods. These values can then be compared to determine if the machine learning solution outperforms these conventional statistical methods, and whether an improvement has been achieved.

16

3. Design & Development

Now the solution is chosen, it must be implemented within the support process. To do so, a default system composition is discussed (3.1). The way the solution can be interconnected within this system composition is examined (3.2). And lastly, ways to anticipate on forecasted anomalies are proposed (3.3).

3.1. Default system composition

Within a default system composition where historical data is collected, there is assumed that three systems exist: The machine which gathers the metrics, a database solution where this data is stored, and a dashboard where the data is visualized. The dashboard retrieves the data from the database and plots it into graphs which the user can interpret for themselves. A schematic overview is available in Figure 16.

Metrics Agent Gathers data	Saves data to →	**Database** Stores data	← Retrieves data from	**Dashboard** Visualizes the data

Figure 16: Default system composition

3.2. Connecting machine learning solution

Now a default system composition is known, the machine learning solution can be implemented. As stated in the requirements in paragraph 2.1, the solution must be able to access training data, store the forecasts, and communicate the forecasts to the user. The solution must therefore be connected to the database where the training data is stored and where it can store the forecasts. Apart from that, there must also be an environment to communicate the forecasts to the user, this is further discussed in 3.3. Figure 17 gives a schematic overview of how the forecasting solution can be implemented.
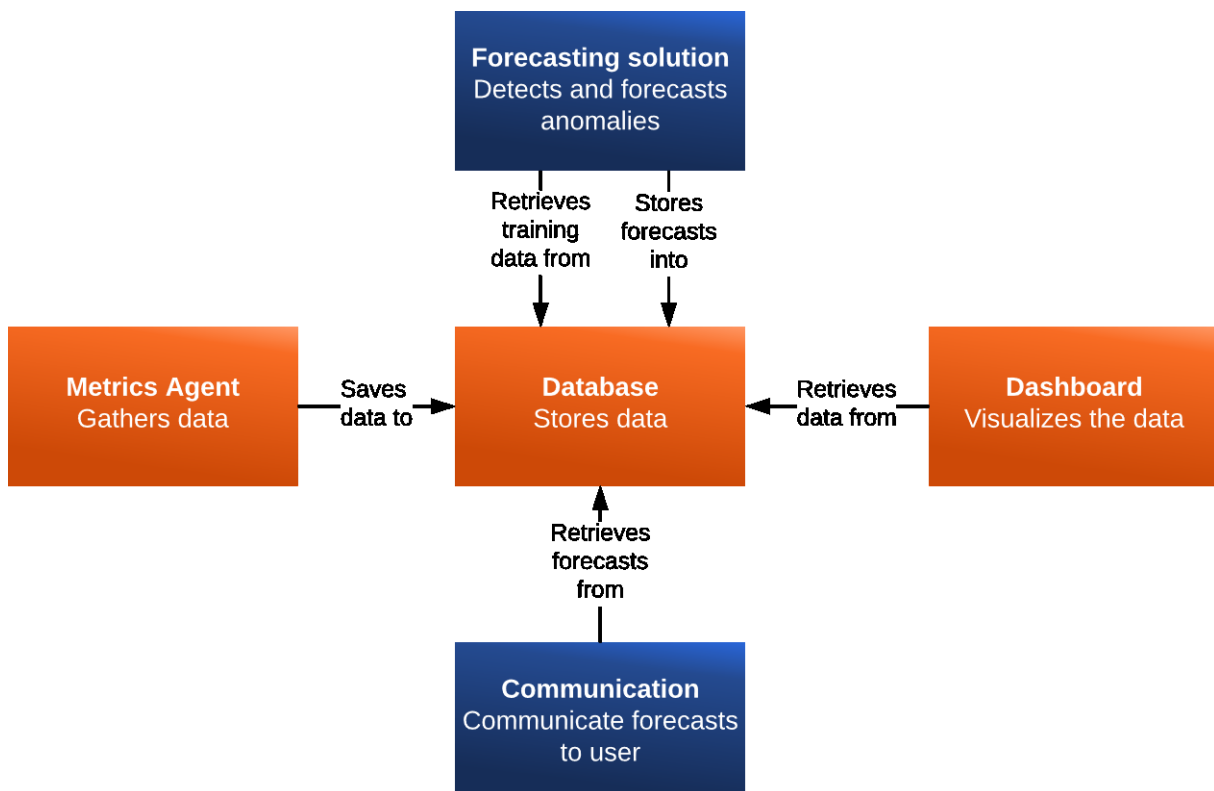
Figure 17: Implementation of solution in support process

3.3.   Forecast anticipation

Now the solution is implemented, the forecasts must be communicated to the company so they can anticipate by making the right business decisions whenever that is required. Propper anticipation prevents anomalies from causing unavailability. To anticipate, the forecasts must be communicated to either the company or a computer which can then make and perform the decision. Therefore, three different communication methods are proposed first, where after a scheme for choosing the respective business decisions is given.

3.3.1. Visualizing on dashboard

The most straightforward and simple communication approach that is proposed in this research is to visualize the forecasts on a dashboard, for simplicity this can be the one which is used to visualize the training data as well. Anomalies in the forecasts can be highlighted and when the user sees one, the user can determine whether to anticipate on the anomaly by making the respective business decision as is discussed in 3.3.4.

3.3.2. Digital or physical alerts

A more preventive approach is setting up digital or physical alerts. When an anomaly is forecasted, the system will send an alert towards a user either digitally, for example via email, or physically, for example by an automatic call. The user is then always aware of the alerts can choose whether to anticipate and how to do so instantly. To realize this, a computer needs to be able to interpret the forecasts and send the email or perform the call. Interpreting the results is a matter of setting thresholds, which are further discussed in 3.3.4. If the forecast exceeds the threshold, a simple computer script, which needs to be developed by the company, has to get triggered and perform the alert. This makes the full implementation more difficult but is more reliable because it is unrealistic to assume that a user is constantly checking the dashboard. The chances of missing an important anomaly are therefore lower than by using the dashboard visualization.

3.3.3. Autonomous responses

The most reliable, but also difficult to implement, is autonomous responses. This is also referred to as self-healing. When an anomaly happens, a machine will adapt on this anomaly by making the required business decision by itself and performing the respective actions. This does require a similar script to be developed as for the alerts, but this script now needs to be aware of what actions must be taken for each of the forecasted metrics, and how to do so. What actions to take is discussed in 3.3.4. Autonomous responses fully relieve the company from the need to have personnel available for anomaly anticipation.

3.3.4. The business decision

What anomalies are necessary to know for the company, as well as the actions which need to be taken when one occurs, differ per metric. For example, companies might only be interested when a specific metric exceeds a maximum value, while for another metric they want to know if it going to reach a value of 0. Therefore, different thresholds must be set for each of the metrics. If a metric is forecasted to reach or exceed these thresholds, an action is required. For a specific metric such an action could mean a system requires a reboot, or for another metric this means a system needs more capacity.

To conclude, for each metric a threshold and a respective action must be defined. So that when a forecast predicts the metric exceeds this threshold an action is done by an employee or script which prevents problems from occurring. To realize this, an addition to the GQM discussed in 2.4 is proposed in Figure 18.

Figure 18: GQM with respective action and example

4. Demonstration

This chapter provides a demonstration of the method by applying it at the commissioning company. This is done in order to create an MVP which can be used for validation purposes. To do so, firstly the current system composition within CAPE, is discussed (4.1). Secondly, the solution is chosen via the method (4.2).

4.1. CAPE's system composition

As discussed in Figure 16, the default system composition of a support process consists of a metric's agent, a database, and a dashboard. The composition of systems at CAPE (available in Figure 25) is expanded way beyond the default composition. At CAPE, they make use of the TICK stack, which provides them with a metric's agent called Telegraf, a database called InfluxDB, and a dashboard called Chronograf (Influxdata, 2019). With this information the suitable machine learning solution can be determined.

4.2. Choosing corresponding machine learning solution

To choose the machine learning solution which fits the situation at CAPE best, the steps which are determined to be taken in 2.3 are used. Firstly, the solutions are limited to those that use the machine learning type which is applicable based on the training data. And secondly, the solution that is most suitable for CAPE from the set that use this learning type is chosen.

4.2.1. Applicable learning type

To determine the applicable learning type, CAPE's trainings data is examined according to the flowchart in Figure 7. Within CAPE no interaction environment is currently available, so reinforcement learning is not an option at this time of writing. Developing one could be considered but as discussed in 2.2.3 the other learning types must first be evaluated to see if they provided sufficient results. Using these learning types takes less effort than to develop a whole new interaction environment. They do have sufficient amounts of training data available, but none of this data is labeled right now. This makes supervised and semi-supervised learning not applicable as well. As discussed in 2.2.3, there can be considered label the data so supervised learning can be used for more accurate results. Because unsupervised learning remains, it is chosen as the machine learning type to be used within this research. If during validation it turns out this results in insufficient forecasts, labeling the data is to be reconsidered in future research.

An additional argument for this decision is the advice determined in preliminary research done within CAPE, Van Tintelen (2019) states: *"Of the two types of machine learning, supervised classification and unsupervised anomaly detection, I would advise the support department to start with unsupervised anomaly detection. The fact that no breakdown data was gatherable makes training of the supervised classification algorithms infeasible. An intensive research in the feature sets before application breakdowns might yield enough breakdown examples. However, as no breakdown examples were found during this research there is no way to guarantee this data will be gathered during a new data investigation. This makes focusing on the supervised classification algorithms a riskier endeavor than to focus on the unsupervised anomaly detection algorithms."*

4.2.2. Matching systems with machine learning solution

To determine which unsupervised machine learning solution to use, the flowchart in Figure 8 is used. As discussed, the training data-source within CAPE is InfluxDB, when searching for a solution with stated compatibility, one arises: Loud ML. As stated on their website:

Figure 19: Loud ML logo (Loud ML team, Loud ML Reference: Getting Started, 2019)

"Loud ML is the first open source deep learning API that makes it simple to prepare, train, and deploy machine learning models and crunch the data stored in your favorite databases without moving the data. The user selects the times series that they want to model and sets the model date ranges, then Loud ML will build the models and save them for inference in production." (Loud ML team, Loud ML Reference: Getting Started, 2019)

Loud ML uses artificial neural networks, in short ANN, for time series forecasting and time series clustering (Loud ML team, Loud ML Reference: Getting Started, 2019). It provides out of the box the Donut unsupervised model. As stated by Xu et al. (2018) Donut was developed with the following purpose: *"To ensure undisrupted business, large Internet companies need to closely monitor various KPIs (e.g., Page Views, number of online users, and number of orders) of its Web applications, to accurately detect anomalies and trigger timely troubleshooting/mitigation. However, anomaly detection for these seasonal KPIs with various patterns and data quality has been a great challenge, especially without labels."*

Xu et al. (2018) describe Donut as: *"An unsupervised anomaly detection algorithm based on VAE. Thanks to a few of our key techniques, Donut greatly outperforms a state-of-arts supervised ensemble approach and a baseline VAE approach, and its best F-scores range from 0.75 to 0.9 for the studied KPIs from a top global Internet company. We come up with a novel KDE interpretation of reconstruction for Donut, making it the first VAE-based anomaly detection algorithm with solid theoretical explanation."*

Based on the proposed method, this solution is thus chosen to be implemented within the process of CAPE. Apart from this engine, a Chronograf version to support Loud ML has been developed as well (Loud ML team, loudml/chronograf, 2019). As CAPE is already familiar with Chronograf, this is used as communication method. Apart from visualizing the forecasts, this Chronograf version provides a graphical user interface for setting up models which is more user friendly than having to set up these models by hand (Loud ML team, loudml/chronograf, 2019). The proposed setup is visualized in Figure 26 (Appendix 1: CAPE's System Composition).
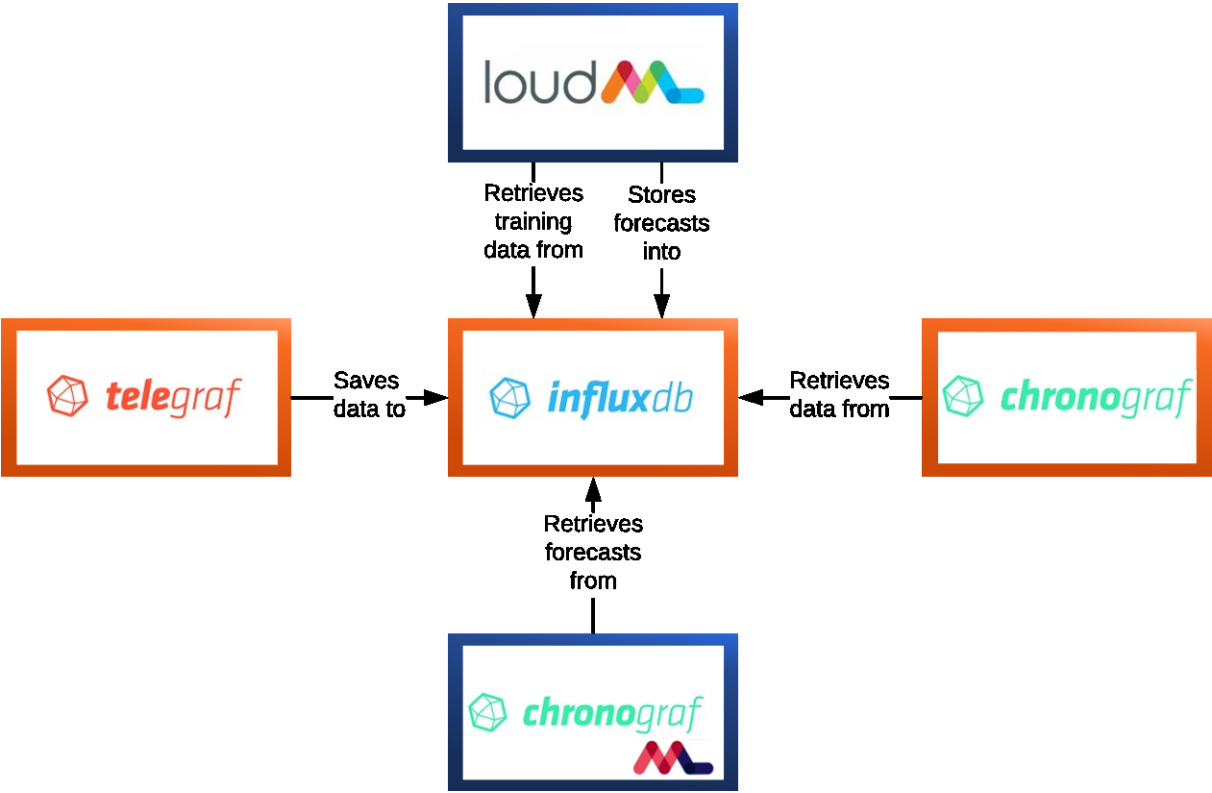

Figure 20: Proposed system composition for CAPE

Now the proposed setup is stated, it can be implemented within the system composition at CAPE. A schematic overview of the full implementation is available in Figure 26 (Appendix 1: CAPE's System Composition).

4.2.3. CAPE's business decisions

Now the environment to forecast metrics is realized. As discussed in 3.3.4, it is time to determine appropriate business decisions for when a specific goal cannot be met due to a metric which will exceed its threshold. As discussed in 1.2.1, CAPE aims to deliver the highest quality for the customers by providing the highest amount of availability to their applications. This results in the goals, qualities, metrics, and actions as visualized in Figure 21. These are to be further extended with more qualities and metrics in the future.
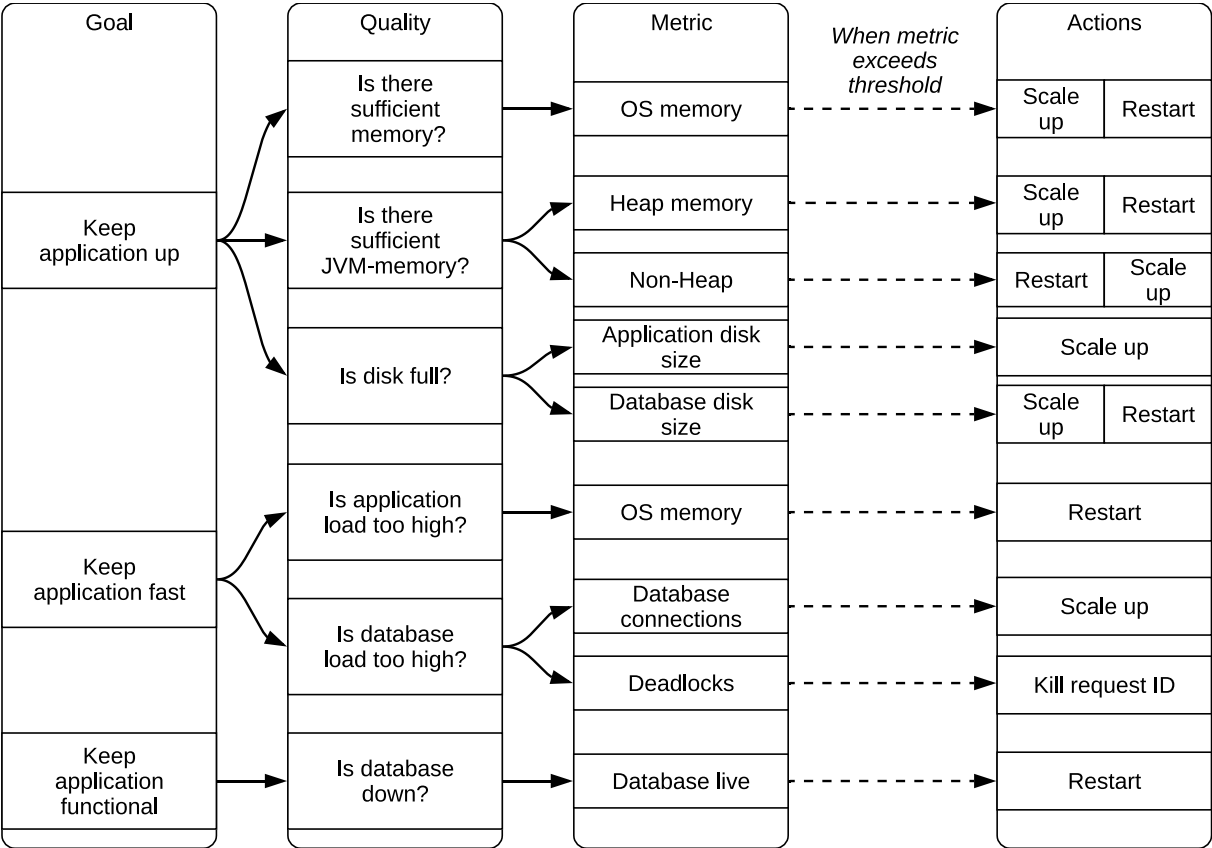

Figure 21: Goals Qualities Metrics and Actions

5. Evaluation

This chapter contains the results of the case validations (5.1), the result of the conventional statistical methods (5.2), and a discussion on these results (5.3). As stated in paragraph 2.5, the method itself is validated based on two objectives. The first one was for the method to be consistent with prior research and was already confirmed to be accomplished there. For the second objective, it must be effective for the intended purpose of 3 cases. The performance of the forecasts is also compared to one of a convention statistical method as discussed in 1.2.5.

5.1. Case validations

Now the machine learning engine is implemented within the system at CAPE, it is used to confirm this second objective, by performing 3 case validations. The chosen cases, based on preferences from employees at CAPE, are:

- Free JVM memory   in the CAPE Service Point acceptance environment
- Free JVM memory   in the CAPE Service Point production environment
- JVM total threads   in the CAPE Service Point production environment

This is done by taking the datapoints for both the forecast as well as the real data and using excel to calculate the measures as discussed in 2.5.

5.1.1. Free JVM memory in acceptance

The forecast for this case is graphed with the actual values in Figure 22. The resulting measures for this case are available in Table 2. The following parameters are used:

- Training period: 2 weeks   (total available training data at time of training)
- Validation period: 24 hours   (available data not used for training at time of validation)
- Interval: 1 minute   (equal to amount of available datapoints)
- Fill: Previous   (to fill up empty gaps with semi-realistic values)

Figure 22: Visualization of case 1 with actual values in blue and forecast in purple

The bias of the forecasts of this case turned out to be $17.01\ MB$ which implies that on average the forecasts are $17.01\ MB$ too low.

Measure	Desired	Reality	Valid
MAPE	10 %	8.10 %	Yes
MAE	200 MB	167 MB	Yes
MSE	40000 MB^2	38106 MB^2	Yes
RMSE	200 MB	195 MB	Yes

Table 2: Results evaluation measures case 1

5.1.2. Free JVM memory in production

The resulting measures for this case are available in Table 2. For this case the following parameters are used:

- Training period: 2 weeks   (total available training data at time of training)
- Validation period: 24 hours   (available data not used for training at time of validation)
- Interval: 1 minute   (equal to amount of available datapoints)
- Fill: Previous   (to fill up empty gaps with semi-realistic values)

The bias of the forecasts of this case turned out to be $52.36\ MB$ which implies that on average the forecasts are $52.63\ MB$ too low.

Measure	Desired	Reality	Valid
MAPE	10%	20.42 %	No
MAE	200 MB	84.64 MB	Yes
MSE	40000 MB^2	10564.45 MB^2	Yes
RMSE	200 MB	102.78 MB	Yes

Table 3: Results evaluation measures case 2

Figure 23: Visualization of case 2 with actual values in blue and forecast in purple

5.1.3. JVM Total Threads in production

The resulting measures for this case are available in Table 2. For this case the following parameters are used:

- Training period: 2 weeks   (total available training data at time of training)
- Validation period: 24 hours   (available data not used for training at time of validation)
- Interval: 1 minute   (equal to amount of available datapoints)
- Fill: Previous   (to fill up empty gaps with semi-realistic values)

The bias of the forecasts of this case turned out to be -7.36 threads on average the forecasts are 7.36 threads too high.

Measure	Desired	Reality	Valid
MAPE	10 %	8.81 %	Yes
MAE	20 threads	11 threads	Yes
MSE	400 threads	229 threads^2	Yes
RMSE	20 threads	15 threads	Yes

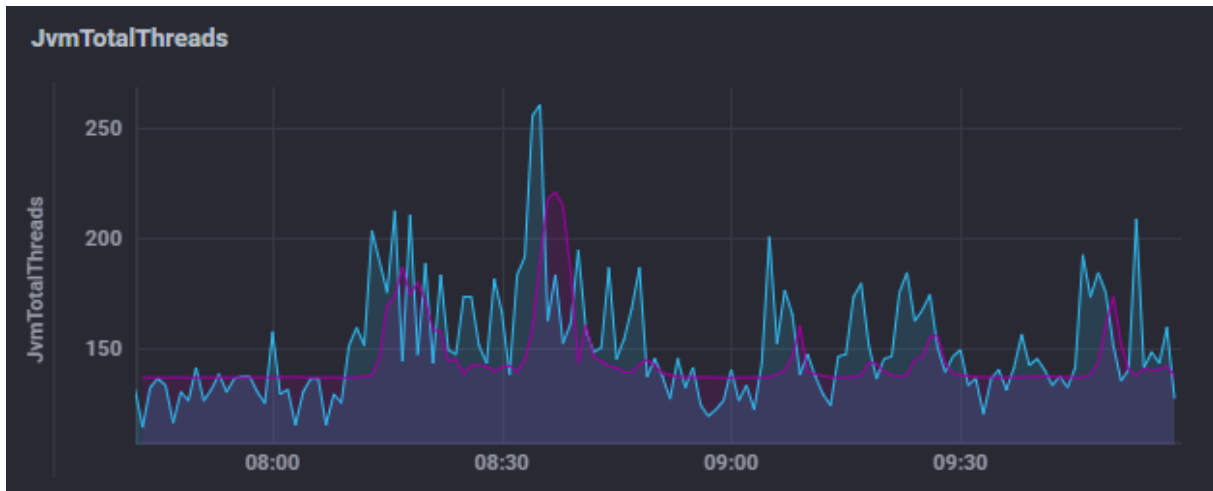*Table 4: Results evaluation measures case 3*

Figure 24: Visualization of case 3 with actual values in blue and forecast in purple

5.2. Conventional statistical method

The forecasts are now done via the conventional statistical methods discussed in 2.5.3. These are validated in the same manner as the forecasts of the implemented solution. These results for each of the cases are in the following tables:

Measure	Loud ML	Moving Average	Exponential Smoothing
MAPE	8.10 %	2.58 %	3.40 %
MAE	167 MB	57.28 MB	75.05 MB
MSE	38106 MB^2	11419.76 MB^2	22564.71 MB^2
RMSE	195 MB	106.86 MB	150.22 MB
Bias	52.47 MB	-0.196 MB	-0.275 MB

Table 5: Evaluation measures Loud ML versus conventional statistical methods case 1

Measure	Loud ML	Moving Average	Exponential Smoothing
MAPE	20.42 %	14.19 %	22.31 %
MAE	84.64 MB	58.78 MB	92.42 MB
MSE	10564.45 MB^2	5168.48 MB^2	12368.83 MB^2
RMSE	102.78 MB	71.89 MB	111.21 MB
Bias	16.93 MB	-0.50 MB	-0.73 MB

Table 6: Evaluation measures Loud ML versus conventional statistical methods case 2

Measure	Loud ML	Moving Average	Exponential Smoothing
MAPE	8.81 %	4.90 %	7.19 %
MAE	11 threads	7 threads	10 threads
MSE	229 threads^2	73 threads^2	160 threads^2
RMSE	15 threads	9 threads	13 threads
Bias	-7 threads	-0.01 threads	-0.01 threads

Table 7: Evaluation measures Loud ML versus conventional statistical methods case 3

As visible in the above tables, the forecasts of Loud ML are currently being outperformed by the ones forecasted with moving average. Moving average is the best for all three cases. Loud ML is second for case 2 and third for the others.

5.3. Discussion

Now the results are in, there can be stated whether the solution performs sufficient or not for both the case studies as the conventional statistical methods.

When looking at the case studies, the solution meets the desired accuracy stated by the company for all three cases, with the exception of the MAPE measure in case 2. This is mainly caused by the fact that the amount of free memory in this production environment is lower than in the acceptance environment, which results in a higher MAPE value while performing similar to the one in case 1 which MAPE is stated to be valid. All three forecasts do show to have a significant bias. Therefore, with respect to the case studies, the models are precise enough but there can be considered to examine how to lower the bias.

With respect to the conventional statistical methods, the forecasts done by the solutions are outperformed in all three cases, especially by moving average. This implies that the conventional methods are more accurate and using Loud ML with the current setup at this time of writing will not provide an improvement over these statistical methods with respect to accuracy. However, the accuracy can be improved when adapting the chosen training model parameters, training data, and forecasting periods. On top of that, Loud ML does provide a graphical environment and API's which enable the company to perform, without effort, a high number of forecasts on all desired metrics for each of the applications maintained by CAPE. When performing the statistical method validations for this research, this turned out to be more difficult and time consuming.

Therefore, the choice for Loud ML is considered to be an important step in the right direction. The accuracy, and especially bias, can be improved and future research about how to improve the training of the models will be beneficial. Besides that, as this research aims to forecast anomalies, it does not validate the business decisions yet. This implies that it is still uncertain if performing the determined actions for the metrics does result in a higher accessibility. Validating these decisions in future research is thought to be beneficial as well.

6.   Conclusion & Recommendations

Now the method is validated it is time to draw a conclusion (6.1) and give recommendations regarding usage (6.2) and possible future research (6.3).

6.1.   Conclusion

This research aimed to develop a method for implementing anomaly forecasting within support processes at companies. Based on a developed minimum viable product there is determined that by following this method, valid forecasts for metrics can be done, which can be used to prognose anomalies and enable companies to anticipate on these before they can cause damage to the company.

To sum up, anomalies can be forecasted within a support process to improve availability by the use of machine learning. For this, a specific learning type can be determined based on the company's data. To execute this type a corresponding solution is chosen according to the data source a company uses. It can be implemented by connecting it to the support process systems and validated to ensure it provides accurate forecasts.

As of now, the method is ready to be used within companies who aspire to implement anomaly forecasting themselves. Some recommendations are suggested in paragraph 6.2 which can improve the accuracy of forecasts by providing standards with respect to the forecasting parameters which are used to setup models. But, even without those, companies can still create accurate models by altering them based on the determined validation method.

This method has further clarified, streamlined, and simplified the implementation process and is hoped to be of great use for many more companies.

6.2.   Recommendations

Based on the above conclusion, the method is ready to be used and all companies with support processes are recommended to implement anomaly forecasting within their own systems, and this method provides a valid way to do so.

Not only can this forecasting be used to improve the provided quality for a company's clients by preventing anomalies, it can also become a delivered service one its own. The number of companies who will want to start using machine learning and forecasting will further increase and starting to sell these as a service is ought to have high market potential.

For now, companies are recommended to base the amount of training data they use, and the length of the periods they forecast on the "Mean Square Error (MSE)" which is determined during the validation. As long as this is within the stated norm, the models are useable and can provide valuable information about metrics and upcoming anomalies.

Companies are recommended to only start forecasting metrics which they have determined to be necessary via the "Goal Quality Metric" approach. It is important they limit the amount of metrics to those to prevent overload. For each of these metrics, the interval between forecasts and the maximum MSE must be determined based on the company their needs.

In paragraph 6.3, future research is suggested to determine more justified methods to determine the required metrics and corresponding parameters.

6.3. Future research

To further improve the accuracy and practicability of the method, the following future research is suggested:

- Additional research can be done about the forecasting model's parameters to increase the accuracy of the forecasts. At this point, no rule of thumb is available about what amount of training data is needed for what amount forecasting. Neither is about the number of anomalies this training data must contain compared to normal behaving data. This will not only benefit this method but the whole field of machine learning as a whole.
- Additional research can be done about whether performing the decided actions do increase the availability of the systems and benefit the company. This validation part was not performed within this research.

Bibliography

Basili, V. R., Caldiera, G., & Rombach, H. D. (1994). *The Goal Question Metric Approach.*

BusinessDictionary. (2019). *forecasting*. Retrieved from BusinessDictionary: http://www.businessdictionary.com/definition/forecasting.html

CAPE Groep B.V. (2019). Retrieved October 3, 2019, from CAPE Groep: https://www.capegroep.nl/

Case, R. (2019, October 27). *Machine Learning, Part 4: Reinforcement Learning*. Retrieved from TowardsDataScience: https://towardsdatascience.com/machine-learning-part-4-reinforcement-learning-43070cbd83ab

Chowdhury, M., & Sadek, A. W. (2012). *Advantages and Limitations of Artificial Intelligence.*

Drakos, G. (2018, August 26). *How to select the Right Evaluation Metric for Machine Learning Models: Part 1 Regression Metrics*. Retrieved from Towards Data Science: https://towardsdatascience.com/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-1-regrression-metrics-3606e25beae0

Galvas, G. (2016). *Time series forecasting used for real-time.* Amsterdam.

Grotov, A. (2019). *The 4 types of data to train Machine Learning models for IT Operations*. Retrieved from StackState: https://blog.stackstate.com/the-4-types-of-data-to-train-machine-learning-models-for-it-operations

Haowen Xu, W. C. (2018). *Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications.* Peking: Tsinghua University.

Heerkens, H. W.-W. (2017). *Solving Managerial Problems Systematically.*

Heidemann, L. (2018, January 18). *Why Anomaly Detection Matters to (Almost) Everyone*. Retrieved from data iku: https://blog.dataiku.com/why-anomaly-detection-matters-to-almost-everyone

Influxdata. (2019). *InfluxDB 1.X: Open Source Time Series Platform | InfluxData*. Retrieved from Influxdata: https://www.influxdata.com/time-series-platform/

Investopedia. (2019, Augustus 8). *Forecasting Definition*. Retrieved from Investopedia: https://www.investopedia.com/terms/f/forecasting.asp

Lehmann, E. L., & Casella, G. (1998). *Theory of Point Estimation.* New York: Springer.

Loud ML team. (2019). *Loud ML Reference: Getting Started*. Retrieved from LoudML: https://loudml.io/en/loudml/reference/current/getting-started.html

Loud ML team. (2019). *loudml/chronograf*. Retrieved from Dockerhub: https://hub.docker.com/r/loudml/chronograf/

Nau, R. (2019). *Duke University*. Retrieved from Moving average and exponential smoothing models: https://people.duke.edu/~rnau/411avg.htm

Nicholson, C. (2019). *Artificial Intelligence (AI) vs. Machine Learning vs. Deep Learning*. Retrieved from Skymind: https://skymind.ai/wiki/ai-vs-machine-learning-vs-deep-learning

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). *A Design Science Research Methodology.*

Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). *Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods.*

Ray, S. (2015, December 29). *8 Proven Ways for improving the "Accuracy" of a Machine Learning Model*. Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2015/12/improve-machine-learning-results/

Rouse, M. (2019). *machine learning (ML)*. Retrieved from TechTarget: https://searchenterpriseai.techtarget.com/definition/machine-learning-ML

Stellwagen, E. (2019). *Forecasting 101: A Guide to Forecast Error Measurement Statistics and How to Use Them*. Retrieved from Forecast Pro: https://www.forecastpro.com/Trends/forecasting101August2011.html

Stewart, M. (2019, March 25). *The Actual Difference Between Statistics and Machine Learning*. Retrieved from Towards Data Science: https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3

Swalin, A. (2018, April 7). *Choosing the Right Metric for Evaluating Machine Learning Models*. Retrieved from Medium: https://medium.com/usf-msds/choosing-the-right-metric-for-machine-learning-models-part-1-a99d7d7414e4?

Tanner, J. (2017, May 24). *Agile Metrics: A GQM Approach*. Retrieved from LeadingAgile: https://www.leadingagile.com/2017/05/agile-metrics-gqm-approach/

Teboul, W. (2018, July 20). *Why use Machine Learning Instead of Traditional Statistics?* Retrieved from TowardsDataScience: https://towardsdatascience.com/why-use-machine-learning-instead-of-traditional-statistics-334c2213700a

Tofallis, C. (2015). A better measure of relative prediction accuracy. *Journal of the Operational Research Society*.

Touger, E. (2018, August 3). *What's the Difference Between Artificial Intelligence (AI), Machine Learning, and Deep Learning?* Retrieved from Prowess: https://www.prowesscorp.com/whats-the-difference-between-artificial-intelligence-ai-machine-learning-and-deep-learning/

Van Tintelen, B. (2019). *Designing a Machine Learning Decision Tree for Information Systems: A study into the implementation of supervised and unsupervised machine learning methods.* Enschede: University of Twente.

Vandeput, N. (2019, July 5). *Forecast KPI: RMSE, MAE, MAPE & Bias*. Retrieved from Medium: https://medium.com/analytics-vidhya/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

Zheng, S. (2019). Methods of Evaluating Estimators. In S. Zheng, *Math 541: Statistical Theory II.* Springfield, Missouri.

Zhu, X. (2005). *Semi-Supervised Learning Literature Survey.* Madison. Retrieved from GeeksforGeeks: https://www.geeksforgeeks.org/ml-semi-supervised-learning/

Appendix 1: CAPE's System Composition

This appendix is **available on the following two pages** and contains the current full system composition of CAPE (Figure 25) and with machine learning implemented (Figure 26).
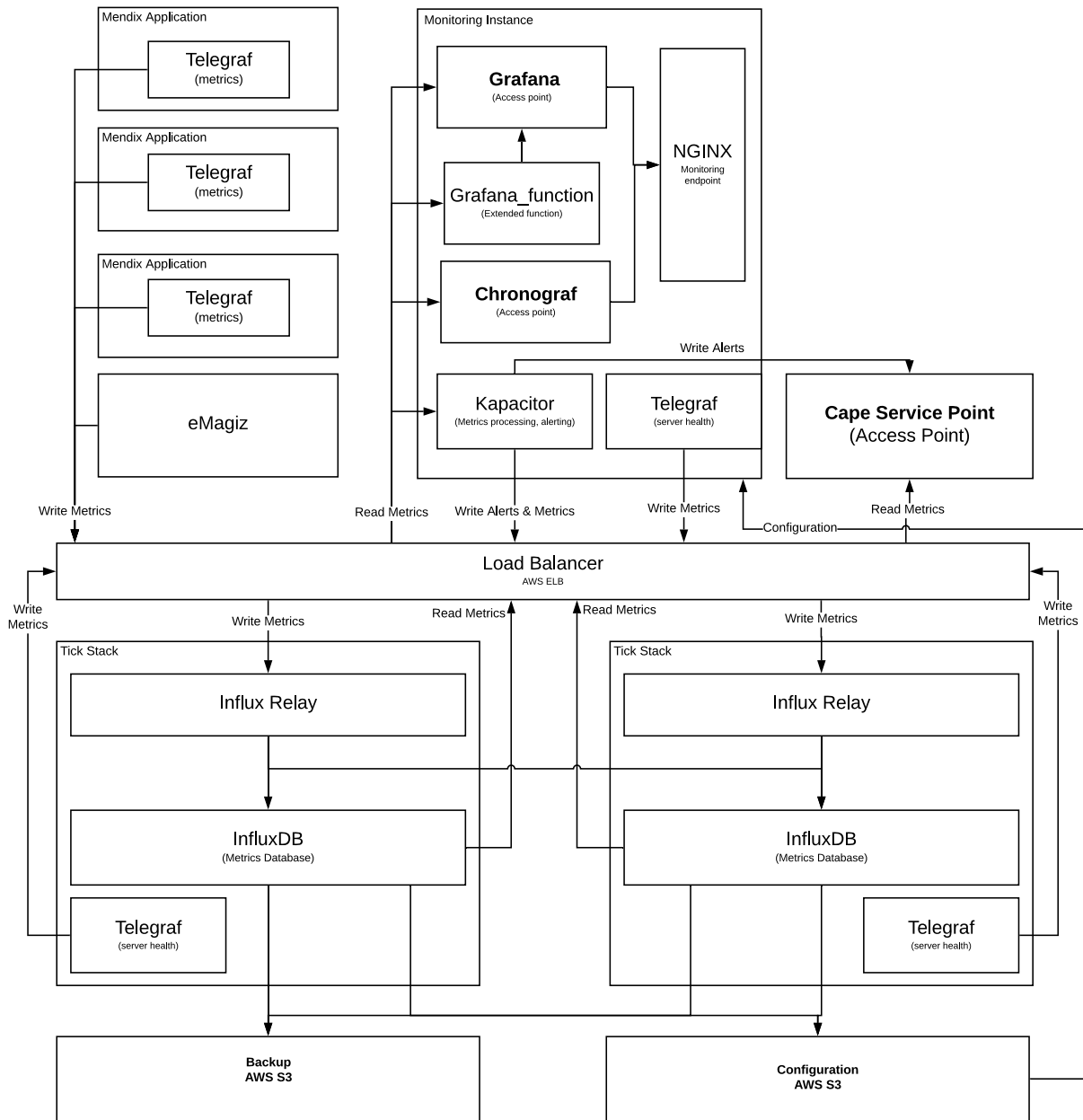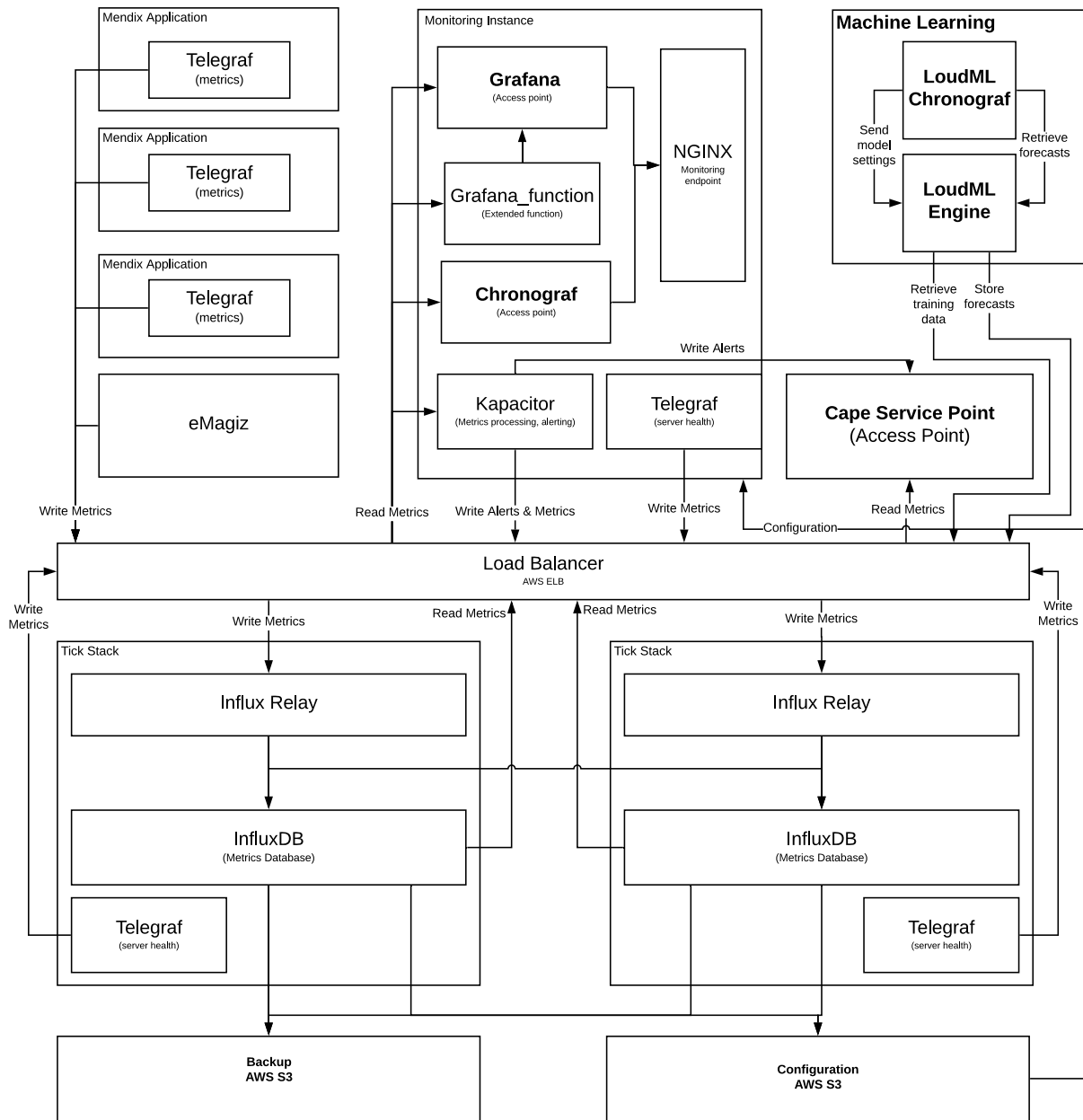
Figure 25: Current system composition at CAPE

Figure 26: CAPE's system composition with implemented machine learning

Appendix 2: Single Page Method Overview

This appendix is **available on the following two pages** provides an overview of the method as a whole.

Step 1: Choose machine learning type based on your training data

Find which type of data is gathered within the company and use the following flowchart to match it with a machine learning type:
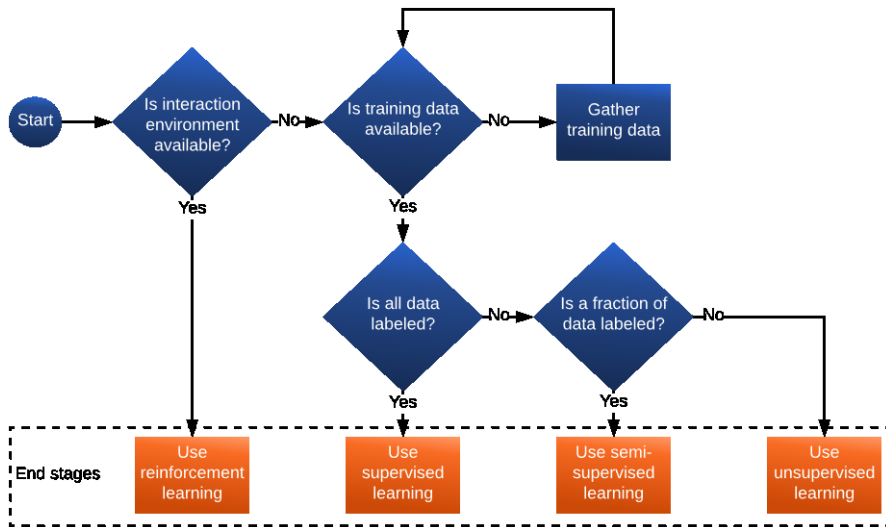

Step 2: Choose respective machine learning solution

Determine where your training data is stored and use the following flowchart to determine a machine learning solution:
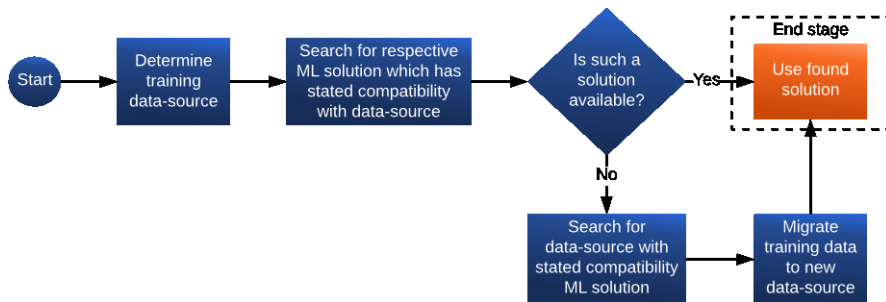
Step 3: Choose metrics to forecast based on GQM approach

Find which metrics are worth forecasting by means of the Goal Quality Metric approach, visualized in the following model:

Step 4: Choose evaluation measure for validation purposes

Choose how to evaluate validity of the model by picking an evaluation measure by using the following flowchart:
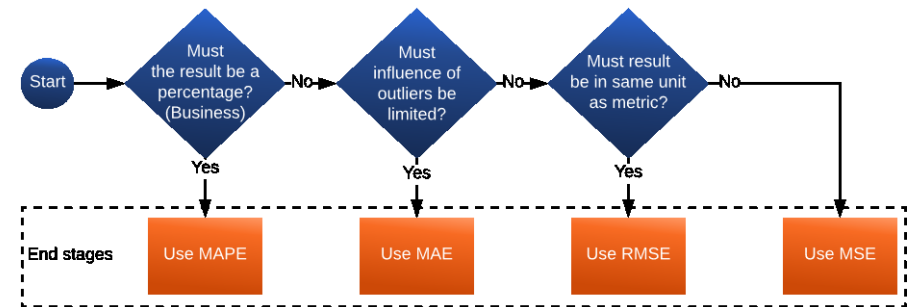
Now state a maximum for this evaluation metric according to the norm you aim to realize.

Step 5: Set business decisions

Determine actions which need to be performed whenever the threshold is exceeded for each of the metrics determined in step 3. Pick one of the following ways to perform these actions and to anticipate on the anomalies:

1. Visualize the results on a dashboard: User interprets anomalies by analyzing the graphs.
2. Create digital or physical alerts: Make a script for example send an email or a phone call to the user with the result.
3. Autonomous response: Make a script automatically correct systems which are forecasted to have anomalies. For example, by rebooting the system.

Step 6: Implement forecasting solution and communication method

Implement the forecasting solution and communication (blue) within the system composition as visualized in the following model. The default composition (orange) consists of a metrics agent, a database where your training data is stored, and a dashboard for visualization of the data.
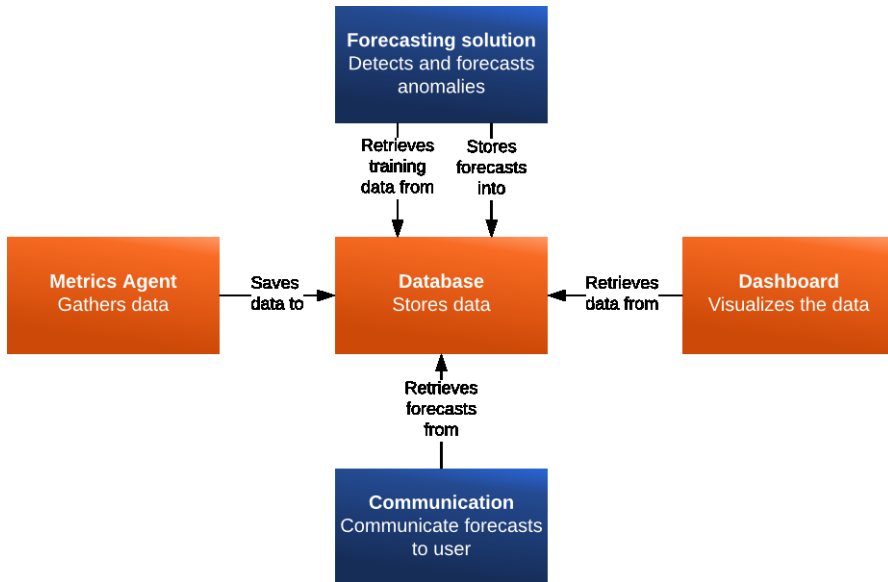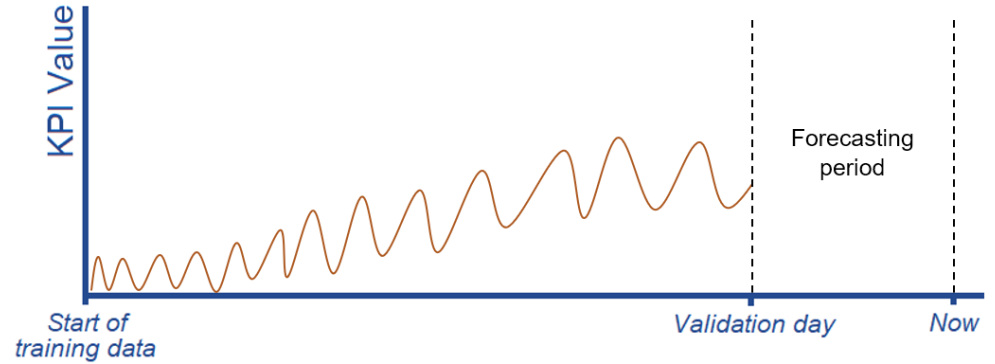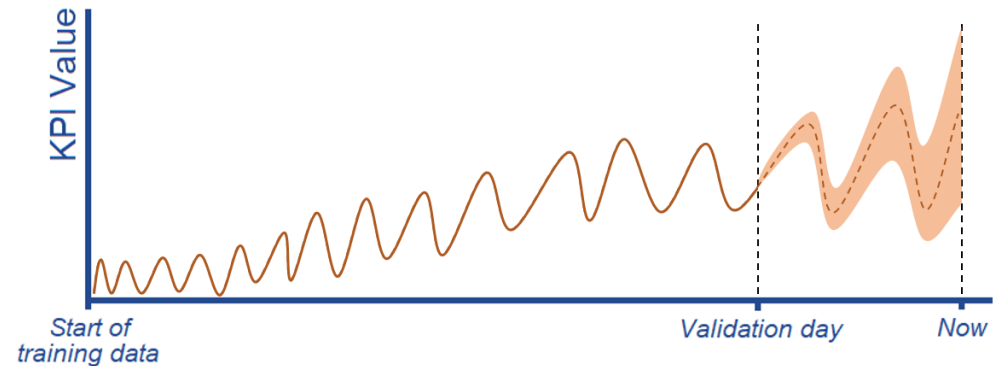

Step 7: Validate the forecasting solution

Retrieve data for comparison by choosing a metric like the following:
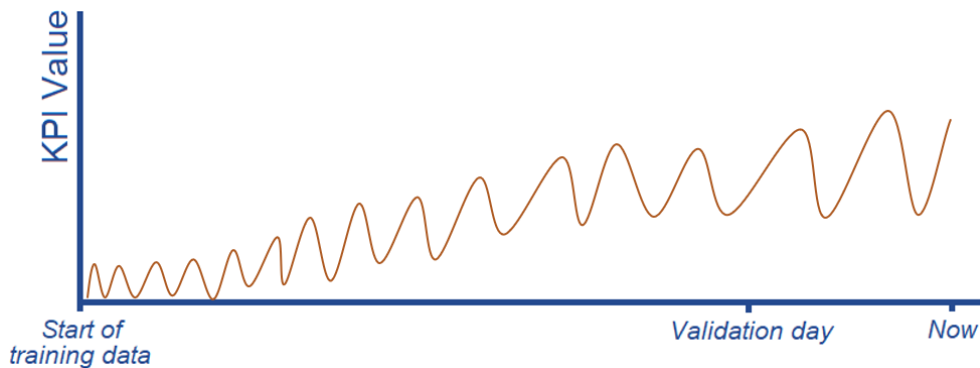
Choose a forecasting period for which you have actual data available:

Make the solution forecast the metric for that period:

Now for all datapoint an actual (or real) value and a forecasted value are available. Calculate the evaluation measure chosen in step 4 as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| * 100\% \qquad MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2 \qquad RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2}$$

Compare this value to your set maximum. If it does not transcend this then: Congratulations, you have successfully implemented anomaly forecasting!

39