LUT University

Software Engineering

Master's Program in Software Engineering and Digital Transformation

Master’s thesis

Nidal Abu-Raed

Predictive maintenance Industry 4.0: case Nokia

Examiners: Professor Ajantha Dahanayake
Jiri Musto M.Sc.

Supervisor(s): Maria Camacho

ABSTRACT

LUT University
Software Engineering
Master's Program in Software Engineering and Digital Transformation

Nidal Abu-Raed

Predictive maintenance Industry 4.0: case Nokia

Master’s thesis

2021

47 pages, 12 pictures, 2 table

Examiners: Professor Ajantha Dahanayake
Jiri Musto M.Sc.

Keywords: predictive maintenance, anomaly detection, cloud computing, Nokia

Predictive maintenance is leading digital transformation to a highly self-optimized and automated environment for humans and machines to work together. In this research predictive maintenance and anomaly detection issues and advantages are reviewed based literature review and real-life experience report. Predictive maintenance introduces advantages technic-wise as well as business-wise. Anomaly detection solution is applied to Nokia AVA cognitive services platform running on top of cloud. It is observed that a lot of challenges found in the literature relating to anomaly detection are common problems during the implementation. Anomaly detection and monitoring services, in general, are a complex problem which requires a lot of resources and planning. The use of existing anomaly detection techniques is suggested.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto
Tietotekniikan koulutusohjelma
Ohjelmistotekniikan ja digitaalisen muunnoksen maisteriohjelma

Nidal Abu-Raed

Ennakoiva ylläpito teollisuus 4.0: tapaustutkimus Nokia

Diplomityö

47 sivua, 12 kuvaa, 2 taulukko

Tarkastaja(t) : Professori Ajantha Dahanayake, Jiri Musto M.Sc

Hakusanat: ennakoiva ylläpito, poikkeavuudenhavainnointi, pilvilaskenta, Nokia

Ennakoiva ylläpito johtaa digitaalista transformaatiota erittäin optimoituun ja automatisoituun ympäristöön, jossa ihmiset ja koneet voivat työskennellä entistä tehokkaammin yhdessä. Tässä tutkimuksessa ennakoivaa ylläpitoa ja poikkeavuuksien havaitsemista koskevia kysymyksiä ja etuja tarkastellaan kirjallisuuskatsauksen ja tosielämän kokemusraportin perusteella. Ennakoiva ylläpito tuo teknisiä ja liiketoiminnallisia etuja. Poikkeavuuksien havaitsemisratkaisua sovelletaan Nokia AVA kognitiivisten palveluiden pilvialustaan. Työssä havaitaan, että monet poikkeavuuksien havaitsemiseen liittyvät kirjallisuudesta löytyvät haasteet ovat yleisiä ongelmia toteutuksen aikana. Poikkeavuuksien havaitsemis- ja seurantapalvelut ovat yleensä monimutkainen ongelma, joka vaatii paljon resursseja ja suunnittelua. Nykyisten olemassa olevien poikkeavuuksien havaitsemistekniikoiden käyttö suositellaan.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all people who were supporting my work, and especially the following people:

Ajantha Dahanayake, LUT School of Engineering Science

Jiri Musto, M.Sc of Computer Engineering

Maria Camacho

Sampsa Tuokko

Thanks for initiating this work and providing me **highly talented** support throughout the work, without I could not have made it.

Also, thanks to LUT for keeping me onboard this far.

Oulu, June 2021

Nidal Abu-Raed

Table of Contents

ATTACHMENTS

LIST OF SYMBOLS AND ABBREVIATIONS

SaaS	Software as a Service
AI	Artificial Intelligence
ML	Machine learning
ICT	Information and Communication Technologies
IoT	Internet Of Things
SaaS	Software as a Service
PaaS	Platform as a Service
PdM	Predictive maintenance
PM	Preventive maintenance
KPI	Key Performance Indicator
IaaS	Infrastructure-as-a-Service
CSP	Communications Service Provider

1 INTRODUCTION

1.1 Background

The worldwide market of cloud computing is new but is already a professionally researched area among IT (Information Technology) companies and scientific communities. The market has been growing for some time now and is still forecasted to expand with high velocity, as it is becoming the de facto capability of IT companies. According to Gartner, the entire public cloud services are predicted to grow by 17% in 2020 to total 266.4 billion U.S dollars [1]. The leading market segment in cloud computing is Cloud Application Services (SaaS) which is the key enabler for modern AI and ML applications provided over the network. SaaS distribution model is trending in software solutions which require high accessibility and versatility and especially among online data analysis tools [2].

We are living in the era of digitalization where business strategies now emphasize digital transformation. According to the article by Enterprise Project, digital transformation can be viewed as "the integration of digital technology into all areas of a business, fundamentally changing how to operate and deliver value to customers" [3]. Through the adoption of modern cloud-based technology solutions, IT companies are now able to develop more customized offerings for customers, with significantly improved quality characteristics from a business point of view. In the big picture, organizations are also enjoying social, economic, and political benefits, as digital transformation is shifting the entire culture of the society and organizations into more digital.

Digital transformation is one of the key drivers of the fourth industrial revolution (Industry 4.0). According to K. Wang [4], industry 4.0 facilitates the vision of Smart Factory, which is about the adoption of new emerging Information and Communication Technologies (ICTs) for the purpose of supporting the manufacturing processes. This leads to a highly self-optimized and automated environment for humans and machines to work together in real-time. According to Duc Tran Anh et. al. [5], the concept of industry 4.0 manifests the following specialized advantages within its environment: flexibility, low production costs, high availability, cost-effectiveness, increased transparency, and resource-saving. There are a lot of buzzwords around digital transformation,

such as *smart cities, smart factories, smart homes, Cyber-physical systems, Internet of Things (IoT), machine learning, AI, big data, predictive maintenance,* etc.

One good example of cloud solutions is one provided by Nokia solutions; an open source-based cognitive service platform called AVA. The challenge that is promised to be met with these services is the increasing network complexity. New services that have prerequisites of extreme network availability, low latency, and high data throughput (i.e., autonomous cars, remote robotic surgeries) are adopting AVA type of services. Nokia is aiming to overcome this complexity with three essential concepts: analytics, virtualization, and automation [6]. Technically speaking, AVA PaaS (Platform-as-a-Service) is a combination of open-source technologies which create the offering. To mention some key functionalities, the AVA platform consists of a unified analytics engine for big data processing, a web-based notebook for interactive data analytics, a database management system, container orchestration system, and distributed computer cluster manager.

This master's thesis will cover concepts of *predictive* and *preventive maintenance*, how and why companies apply these frameworks within this era of digitalization. The literature review will provide fundamental knowledge about predictive and preventive maintenance, which will be used for the implementation of anomaly detection applications on AVA PaaS. The developed system will manifest the key aspects of these frameworks.

1.2 Reseach goals and delimitations

The goal of this thesis is to introduce the concept of predictive maintenance and anomaly detection. In addition, based on the knowledge, develop an anomaly detection application with AVA cognitive services platform. The research question of the thesis:

- RQ1. What are the impacts and benefits of a well-implemented predictive maintenance strategy from the system point of view? Management point of view? Generic view?

- RQ2. What are the existing techniques and technologies for predictive maintenance?
    - SQ1. What are the key technical elements in utilizing predictive maintenance?
    - SQ2. What are known challenges in anomaly detection?

The literature review in this research is limited to **technical** challenges and elements of anomaly detection. Only research papers that include content directly supporting the implementation of the anomaly detection system are used.

The main delimitation regarding the implementation is the time to develop and integrate the developed application. Developing the application would require a lot of testing which is limited by the contract of employment in Nokia. Therefore, only the MVP (minimum viable product) of the application is in the scope of the implementation.

Another delimitation is that the implementation is focused on open-source applications, and the integration is concentrated on a specific cloud service model (PaaS). Also, the implementation must comply with the tools that exists on the platform.

1.3 Research methods

A systematic approach will be used to carry out the literature review on predictive maintenance and challenges relating to anomaly detection techniques. The literature review process follows the steps presented in Figure 1 below.

Figure 1 Steps towards literature review

Table 1 below provides a list of scientific databases and keywords that will be used for the search.

Table 1 Results and keywords used in scientific databases

Scientific database	Predictive maintenance challenges	Anomaly detection AND challenges
LUT Primo	165	59
IEEE Xplore	4054	1874
Science Direct	~510 000	~34 000
ACM	~321 000	~114 000
Google Scholar	2 800 000	182 000

When searching sources, defined keywords are used in the search. Furthermore, relevant papers should include some of the predefined keywords in the title, abstract, or in the keywords section of the paper. All abstracts must be read carefully.

1.4 Structure of the thesis

In chapter 2, literature about predictive maintenance is reviewed to build a foundation for understanding the concept of this maintenance strategy. This chapter also discusses the main impacts and benefits of predictive maintenance from both the management and maintenance team's points of view. In addition, Nokia's AVA (cloud-based platform for analytics) and its key concepts are introduced in this chapter. The purpose of this chapter is to answer RQ1.

In chapter 3, literature will be explored to determine what techniques and frameworks are existing for the predictive maintenance use cases, and what challenges predictive maintenance applications typically encounter during implementation. The purpose of this chapter is to answer RQ2 and its sub-questions SQ1 and SQ2.

Chapter 4 presents an experience report on Nokia, and the development of the practical implementation of an anomaly detection system, as well as an overview of the experience during the development. In this chapter, the challenges encountered are discussed and how they were reached. Also, the integration process with AVA PaaS is covered. Finally, the key aspects of the system in terms of predictive maintenance are described.

2 PREDICTIVE MAINTENANCE AND AVA

2.1 Predictive maintenance

What is predictive maintenance? In software engineering, there are several types of maintenance modes; corrective maintenance, adaptive maintenance, perfective maintenance, preventive maintenance, reactive maintenance, to mention some. However, preventive, and predictive maintenance is currently the state-of-art in software maintenance. These concepts are currently hyped, and they are the result of Industry 4.0 and digital transformation in general.

K. Wang from Norwegian's University of Science and Technology defines predictive maintenance as follows [4]:

"Predictive maintenance is a set of activities that detect changes in the physical condition of equipment (signs of failure) in order to carry out the appropriate maintenance work for maximizing the service life of equipment without increasing the risk of failure."

Lughofer Edvin et. al. generic definition of predictive maintenance techniques [7]:

"Predictive maintenance techniques are designed to help determine the condition of in-service equipment in order to estimate when maintenance should be performed"?

Another interesting definition of predictive maintenance based on the book called O&M Best Practices Guide Release 3.0 [8]:

“Measurements that detect the onset of system degradation (lower functional state), thereby allowing casual stressors to be eliminated or controlled prior to any significant deterioration in the component physical state. Results indicate current and future functional capability.”

Predictive maintenance is the advanced approach of condition-based maintenance strategy which exploits automated condition monitoring by computerized evaluation of machinery input data and allows smart predictions to identify signs of failure and when they are most likely to occur. The foundation of predictive maintenance is based on the concept of real-time measuring and forecasting of machine states. Meaningful metrics are being collected at regular intervals, and maintenance is scheduled as needed based on asset conditions [9]. These metrics are called Key Performance Indicators (KPIs) and they are at the center of the monitoring process. KPIs are specific metrics that are planned and collected depending on the maintained item. For example, when monitoring the performance of a software platform, the maintenance team could set disk usage, CPU usage, memory usage as KPIs to monitor. The concept of predictive maintenance comes into the picture when KPIs are being predicted for future maintenance activities. Sometimes KPIs are derived from multiple different metrics.

Through the development of machine learning and AI, maintenance teams are enjoying this advantage. Machine learning algorithms enable the forecasting of KPIs using historical data. Typically, ML algorithms are trained with some sample data, and real-time data streams, and this way they improve and learn in real-time. As they evolve, they can produce more accurate results and precisely flag unexpected behavior as data points (anomalies), which are then forwarded to the maintenance team as an alert. From a technical point of view, the forecasting of the decomposition of an item assumes that most anomalies do not happen suddenly, but instead, there is a clear evolution from normality to abnormality. The process of monitoring item behavior is typically supported by visualization such as intelligent dashboards. So, with the modern setup of technologies, the maintenance teams’ responsibility is left to investigate dashboards and rely on automated and scheduled maintenance tasks, and of course maintaining and improving the intelligent maintenance infrastructure.

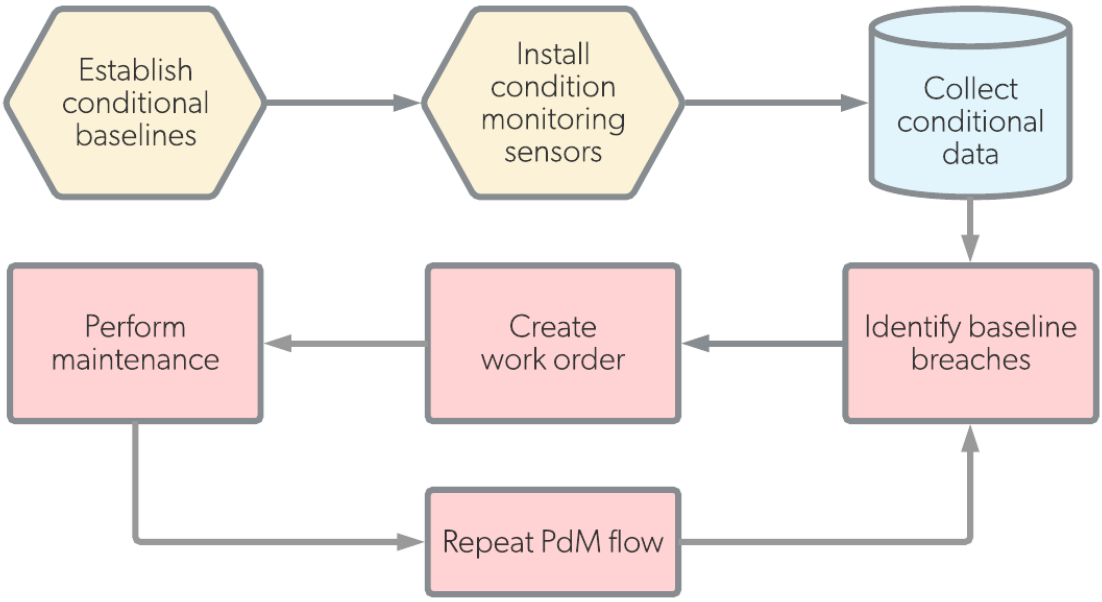


Figure 2 Predictive maintenance workflow (https://www.onupkeep.com/learning/maintenance-types/preventive-predictive)

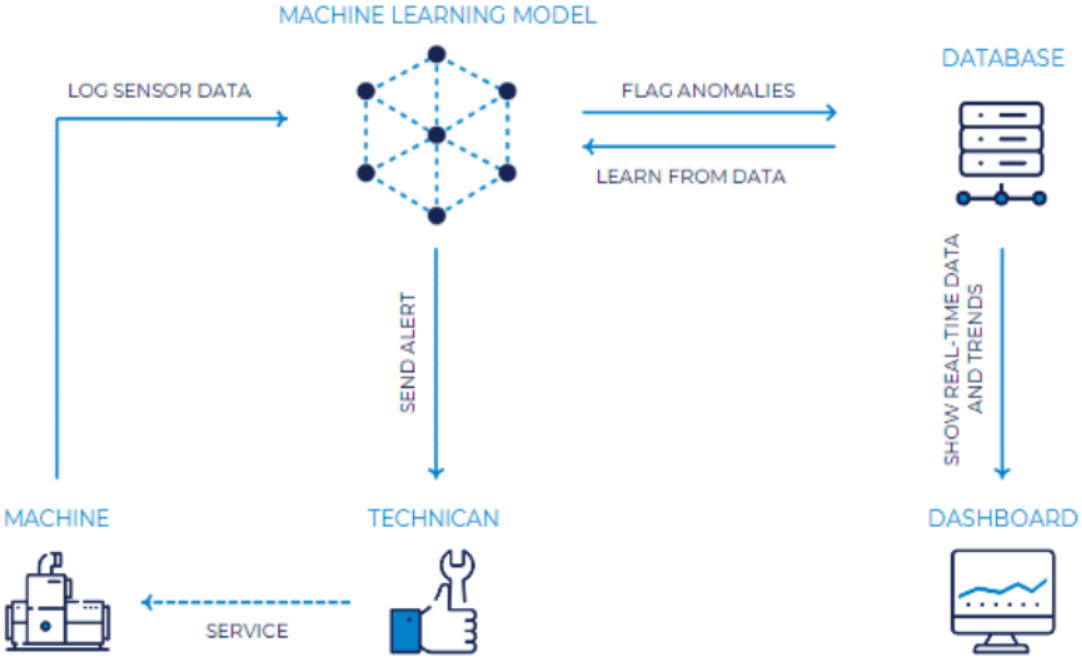


Figure 3 Practical view of PdM in companies (https://theblue.ai/blog/predictive-maintenance-practical-use-in-companies/)

2.1.1 Advantages and challenges

There is a lot of discussion about the advantages that companies enjoy when deploying predictive maintenance programs. Compared to the old times of software maintenance most of the work was done manually, but also in a reactive manner. Meaning that only after a problem has been detected, the solution or fix are applied. Proactive measures can of course be in place, like setting thresholds and alerts that notify a threshold breach. But the usual way of implementing so used to be creation of bash scripts scheduled via cron jobs. In a distributed system, this type of solution is error prone and needs constant maintenance and attention.

In a generic sense, one great advantage is the increased component operational lifetime/availability. Forecasting and extreme automation allow for preemptive corrective actions and protect the system from equipment failures. Also, the lifetime of systems will be enhanced significantly as maintenance teams are able to schedule maintenance activities to optimize the overall usage of the system which saves the system from regular aging of components caused by irregular consumption. This is directly impacting the long-term costs, that the system might have [9].

It has been observed that predictive maintenance also decreases the equipment and process downtime through optimization of operations of equipment [9]. This impacts the overall reliability and availability of a software product.

As mentioned previously, enhanced lifetime and availability, decreased downtimes, and lower overall costs are all together improving software product quality. This leads to better customer loyalty and satisfaction in general. Users are much happier to use software that is doing the job for the long term and is reliable and safe both environmentally and technically [9].

In addition, when automating maintenance and operations tasks, it will release time from developers/workers. This subject has been studied and it seems that adopting predictive maintenance strategy improves productivity and overall effectiveness of the manufacturing environment. It has a positive impact on worker and environment safety, and worker morale as well [9][10]. Furthermore, moving into a more intuitive and automated environment, flexibility

comes into the picture and developers appreciate it. From the managements' point of view, this means that developers will have more time to develop new capabilities (with the support of the company), which are necessities for the new revised roles [11]. Continuous seeking of new opportunities and advanced communication skills are well needed in managing new interdependencies.

From managements' point of view, one major challenge behind this digital transformation is the question that how to re-organize the IT department to fulfill the new job profiles. Research conducted the University of Edwardsville investigates how cloud computing (which is the key enabler of digital transformation and specifically predictive maintenance) is reshaping the role of IT departments. When I am talking about cloud computing, I am referring to the shared pool of computing resources that are provided over the internet. The research claims that cloud computing is causing the traditional IT department to obsolete and decrease its purpose. This article [12] discusses how cloud computing is transforming companies' IT departments statuesque and what IT companies should consider maintaining and improve their overall firm performance while operating in a highly competitive and changing ecosystem. In the past, the IT department of a company was responsible for streaming information to internal customers of a company, but as cloud computing is considered as part of the transformation, it brings a new party to the ecosystem, the cloud vendor. Thus, companies will have to consider how to re-organize their IT department to still bring value for the company, since the cloud vendor itself can be easily coupled with firms' internal customers decreasing the meaning of the IT department.

However, the paper also describes some key advantages and disadvantages of cloud computing and captive IT. Most of the disadvantages of traditional captive IT departments have been cost-related things like high maintenance costs and infrastructure investments and the advantages that are considered are mostly just perceived advantages. Cloud computing is seen to abolish these disadvantages totally and amplify the advantages in a more diverse manner. The essential disadvantage that comes with cloud computing, is that the cloud vendors' objectives might not be aligned with the customers' objectives; this establishes the sustained reliance on the IT department. However, this requires some changes in the roles of the IT department. The paper argues that there are two different roles that the IT department must transform into external-facing and internal-facing. This kind of transformation requires the development of new capabilities within the

company and re-organizing of resources. IT departments must have ongoing processes to evaluate, procure, bill, and monitor different cloud vendors (external-facing), and provide new managed services for internal customers with the help of BA and BI (internal-facing), for example. IT departments need to couple with internal customers, to understand the business needs and objectives to create more.
firm-specific opportunities. To achieve this, firms must create new organizational mechanisms that enable close collaboration between IT department leaders and business managers [12].

The paper also discusses the importance of evaluating certain service attributes of cloud-based services to identify efficient and quality services among different cloud vendors, and so to improve the firm's overall performance. In addition, the author has emphasized several times the importance of measuring firms' performance through accounting measures. Firms' performance is considered to increase by following all the previously mentioned disciplines [12].

A literature review conducted by Jovani D. et. al. [13] Includes a comprehensive set of challenges that has been studied in the literature about machine learning in the context of predictive maintenance. The analysis shows that generic challenges relate to the reduction of maintenance-related costs or aims at improving productions efficiency. These challenges are divided into three groups: big data analytics, machine learning models and ontology.

Within the domain of bid data analytics, some issues are concerning the need for real-time monitoring and processing of a large amount of data. In this context, in order to ensure good values on metrics, the challenging areas are latency [14, 15, 16], scalability [14, 17. 18], and network bandwidth [14, 15]. This because some predicted values require immediate action to prevent failures. These challenges, however, are well researched and there are techniques to tackle them for example distributed computing, and edge computing [13]. Other challenges in this domain are data acquisition [17, 19, 20].

Other issues in predictive maintenance in context of machine learning and ontology are described to be, data quality, dynamic environment for operations [13, 21, 22, 23, 24, 25, 26], data understanding, need for specialists [13, 27, 28], evaluation and selection [21], suitability of models,

context awareness [29], interoperability, storing of information semantically and computational cost/time [13, 27].

2.2 AVA Cognitive services platform

AVA is Nokia provided; cloud-based data analytics platform (PaaS) builds from a set of open-source solutions. More specifically, it is a complete AI (Artificial Intelligence) as-a-Service offering, which combines data science, machine learning as well as telecommunications and cloud expertise. Nokia AVA Cognitive Services platform is designed to help companies of telecommunications overcome the new complexities that come along with 5G, and the explosion of the Internet of Things -devices. Communications service providers (CSPs) are facing new critical challenges: complex new use cases, strict service level requirements (low latency, high data throughput, extreme resilience of the network), virtualization, and network slicing. AVA framework provides an end-to-end service view with near real-time impact correlation for better visibility and control, supported by an extensive library of AI use cases to meet these challenges. In general, AVA services aim to improve operational effectiveness and enhance customer experience through extreme automation [30][31]. Figure 3 illustrates the overview of the key concepts and services around AVA.

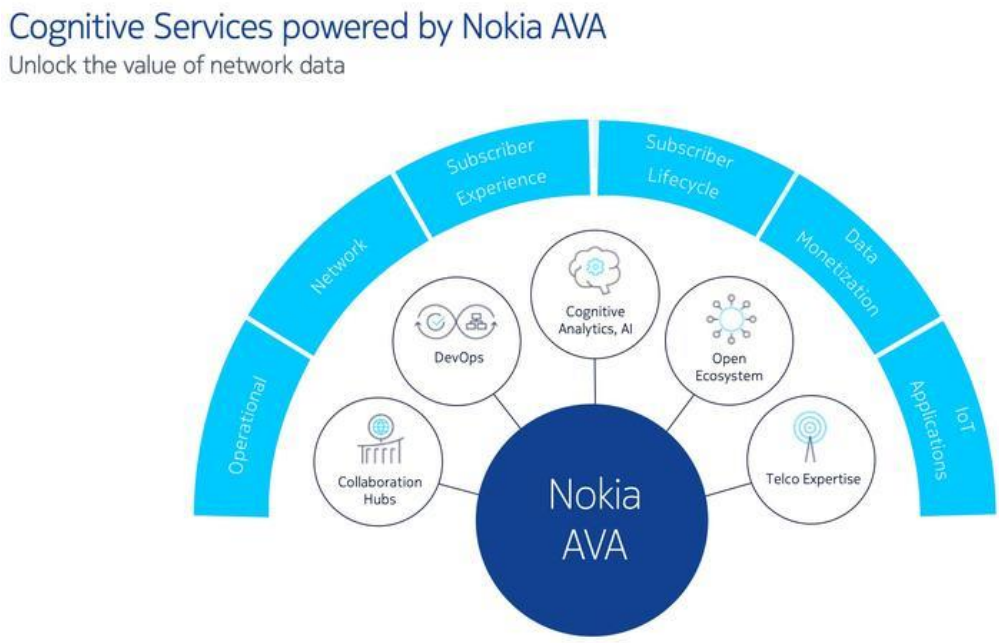


Figure 4 Nokia AVA key concepts and services [32]

The AVA framework consists of analytics, virtualization, and automation technologies. These concepts result in superior network performance for operators [33]. Figure 4 illustrates the conceptual view of the AVA Platform.

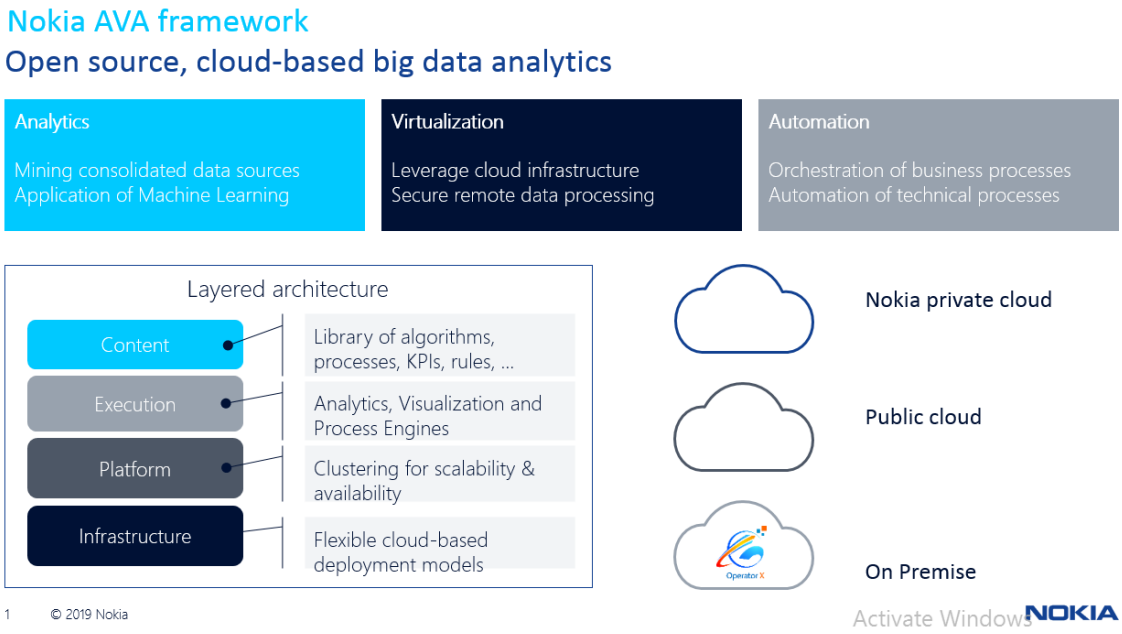


Figure 5 Nokia AVA conceptual diagram

As mentioned before, the architecture of AVA is built as layers of open-source components. The bottom layer of the architecture is IaaS (Infrastructure-as-a-Service) layer provided by Microsoft Azure cloud [34]. This infrastructure layer is the foundation of virtualization, networking, big data storing, and other hardware. One of the major benefits of Microsoft Azure IaaS is that it provides flexible cloud-based deployment models, making it easy and fast to deploy infrastructure with Nokia's private clouds, customer clouds, or on-premises, without restricting the running of AVA PaaS to any specific infrastructure. The AVA platform as a service provides the runtime environment, middleware, and operating system, for developing data analytics applications.

Even though PaaS and IaaS are provided by Microsoft Azure, the solution has been developed in-house adopting open-source components for features like data communication and visualization. Also, the software has some embedded third-party services for specific use cases as well as machine learning algorithms developed by Nokia Bell Labs.

3. ANOMALY DETECTION TECHNICAL CHALLENGES AND ELEMENTS IN LITERATURE

A literature review was conducted to collect information to answer RQ2 and its sub-questions SQ1, and SQ2. The goal of the review was to find information about the existing techniques, technologies, and challenges of real-time anomaly detection in the context of a cloud environment. However, the challenges were researched in the generic sense as well, to get an overall view of the phenomenon. The data was searched from the following databases: LUT Primo, Google Scholar, Science Direct, IEEE, ACM.

The results were examined systematically by inspecting predefined terms used in the titles, reading abstracts carefully, and scouting out the keywords of a paper. Because the research questions are so generic, there was no need to determine any theoretical screening criteria for the papers or to use advanced research methods.

3.1 State of research

To understand the existing research field concerning anomaly detection in the context of the cloud environment, the state of the research was briefly investigated. Nokia's AVA platform has not been under research or gained any attention among research communities, so this study is one of a kind and is going to fit into the surroundings of the research agenda.

Riyaz et al. conducted a comprehensive survey on the state-of-art real-time big data technologies, applications, and existing anomaly detection techniques in 2018. The research explores 13 different studies ranging from 2013 to 2018, which were picked carefully based on the research objectives. Firstly, the objective was to gain insights into real-time big data technologies, applications, and anomaly detection techniques. Secondly, it presents details about the relationships between anomaly detection, machine learning, and real-time big data processing, and the taxonomy of these domains. Then lastly, the research identifies and discusses the challenges of real-time big data processing of anomaly detection [35]. At the end of the paper, Riyaz et al. also discuss possible future research topics.

Based on Riyaz et al.'s paper the current state of research in this field is focused on batching processing instead of real-time processing. However, the concept of batch processing takes an advantage of the same tools and some techniques as real-time processing, so this study is still applicable for this research; it is rather a matter of configuration, and on some part, about the selection of tools. The study also reveals that the research is in at a mature state and there is a lot of real-life studies [35]. Real-world applicability is the key concept of this research since it attempts to serve insights to the challenges and solutions which will be applied in AVA development.

Riyaz et al. propose future research directions on several different technical aspects of real-time anomaly detection (redundancy, nature of input data, parameters selection, inadequate architecture, and many more). In addition, they have scoped these topics into specific development suggestions for the community, and so they addressed that there is a need for developing a more adaptable and responsive model for real-time data processing, which includes a separate module for data labeling and reformatting. This would allow easier run-time processing of data and retraining of detection models in real-time. [35] Thus, AVA development will fit into this proposal very well.

This research is closely related to the above-mentioned research areas, and they will provide the foundation for the AVA development. In the context of AVA and anomaly detection, the studies will provide concrete requirements and challenges for the development.

3.2 General view on anomaly detection

Perhaps the first question that comes to developers' minds when starting the development of anomaly detection applications with cloud, is the question where does the application fit? On which layer of architecture, the application will be developed. Anomaly detection attempts to take advantage of the massive amount of data collected by different KPI collectors (e.g., sensors, embedded monitoring software, internet of things). Figure 5 describes the layers of anomaly detection systems. Anomaly detection being on the top of the diagram indicates that all the other layers below are key **enablers** for high capacity and performing anomaly detection setup. Each layer is an independent area of expertise, and they are developing at their own pace regarding the markets. In other words, as big data processing technologies, network and storage infrastructure, smart environments are developing, the relevance of anomaly detection becomes much greater and demand for it increases.

There are several viewpoints into the benefits of anomaly detection (technical, financial, environmental, management) like machinery lifetime, performance gains, operation costs, energy savings, and agility. However, anomaly detection should bring benefits from any of these perspectives, otherwise, it becomes a useless application for IT and business. Perhaps, the most significant aspect that anomaly detection plays a high role in is **quality** (service, way of working, leadership, costs).

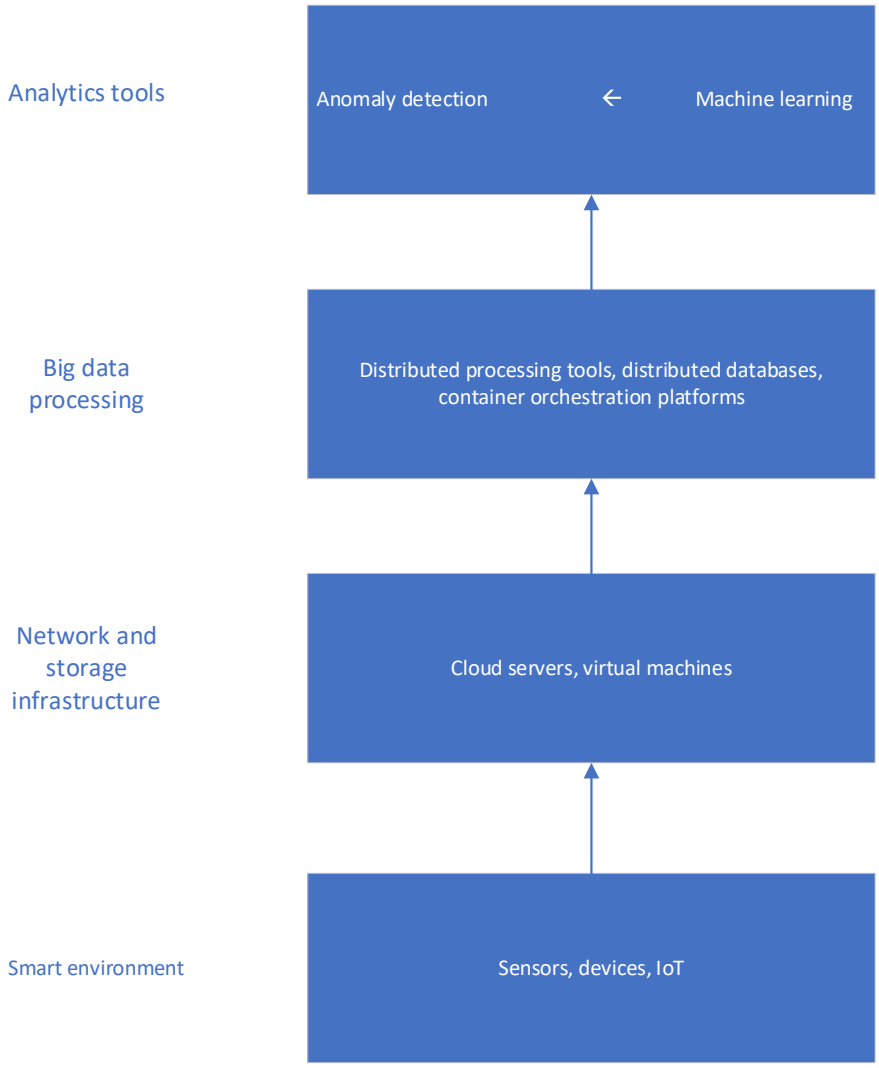


Figure 5 Architecture layers of massive anomaly detection systems

3.3 Anomaly detection challenges

There has been a lot of research done around anomaly detection and real-time big data processing technologies. It seems that there is a lot of common challenges, and studies report that some challenges are typical because of the nature of the problem. A survey conducted by Abdullah G. et. al. [35] reveals common challenges relating to anomaly detection and big data processing technologies. These challenges being:

- Redundancy
- Computational cost
- Nature of input
- Noise and missing values
- Parameter selection
- Inadequate architecture
- Data visualization
- Heterogeneity of data
- Accuracy

Redundancy as a challenge relates to the problem of having a massive amount of data due to the generation of duplicate data, inconsistent data, and the cost of maintaining storage data. The survey suggests that there is demand for a framework that deals with redundancy issues and that the existing tools are not capable to minimize redundancy issues enough [35].

Due to the fact of large sample sizes, the use of anomaly detection techniques and along with cloud can cause heavy computational costs and algorithmic instability. However, there are ways to reduce computational costs, but this requires extra efforts in the planning work of data processing methods/setups [35].

The pain point in input data, is that anomaly detection techniques vary incompatible input data formats. Every anomaly detection technique has a unique way of formatting input data, which brings limitations to the use of the specific techniques. In some cases, there is a need to modify the functionality of data formatting [35] [36].

Noisy and incomplete data are caused by different data transfer speeds and data types – which potentially can cause reliability of anomaly detection. In practice, incomplete data results in false alarms, and missing of true anomalies in machinery behavior [35] [37].

The problem of parameter selection is related to the training of models. It is time-consuming to test a different set of parameters and trying to find the best possible combination of parameters for optimized performance of the anomaly detection model [35]. Also, it is said that high training costs are related to anomaly detection in general [36].

Inadequate architecture is a common challenge in real-time anomaly detection. The survey points out that the existing architecture is capable to manage big data in batches, but not in real-time streaming. Also, the survey reports that organizations are seeking to develop a new architecture for an inadequate architecture of big data processing [35].

The problem of data visualization lies in the network of multiple different devices connected which makes it harder to find an appropriate technique for data visualization. The survey highlights that embedding open-source visualization techniques will result in this problem [35].

The challenge of heterogeneity of data kicks in when unstructured data should be used for anomaly detection because it is not feasible to reformat unstructured data into structured – thus this challenge will cause extra efforts to the data categorization strategy [35].

Based on the survey, accuracy is a widely known issue in the field of anomaly detection, since it is highly linked to all previously mentioned challenges. This challenge is expected to be tackled in the future by improving the current architecture of big data processing. However, some accuracy issues can be compensated with high computational resources which on the other hand, then leads to increased computational costs [35] [36][38].

Defining normal behavior of data from anomalous is challenging. This is because sometimes data might contain complex patterns that appear normal which in fact are anomaly due to malicious actions for example. Also, in some cases data might be corrupted [37].

Also, the report points out the adequate data from machinery failures is hard to obtain because failures do not happen often. This makes it hard for the model to learn from data for future failures.

The report suggests that the anomaly detection approach suits more for health monitoring tasks by monitoring deviations in healthy machinery data [35] [36][39].

From higher point of view, the common challenges describe above can be classified under designing phases of anomaly detectors, which are **model creation**, **model deployment**, and **model retraining and retuning** [40]. In addition to the above-mentioned challenges, Ahmed C. et. al. has identified scalability as challenge as well, due to lack of labeled datasets. Also, they have identified that distribution shift in physical machinery state is a challenge from the retraining and retuning point of view, if a machinery was required to change its processing, then the new generated data would raise alarms from the anomaly detector, however in real world this would be normal behavior. Thus, the model would require more training and tuning to adapt to the new way of working, and the model training would require repetitive work to be done. Distribution shifts are tricky to be observed from the data due to component degradation.

4. EXPERIENCE REPORT ON NOKIA AVA FEATURE DEVELOPMENT

A cloud-based data analytics platform development project was run by Nokia during the time of this work, the year 2019. The platform applications – mostly open-source software applications – were under development and configuration management. This experience report presents the fundamental steps taken to integrate new open-source functionality into the platform/project.

At the time of the integration, the project was completed without the knowledge of the challenges in the literature nor background knowledge on predictive maintenance as described in the previous chapters. The solutions are now cross analyzed with the findings presented in chapter 3 as a post-observation study.

In the scope of this experience report, the report concerns both technical and development operations-related issues.

4.1 Background

The Nokia AVA platform was ramped up and maintained almost 3.5 years before the application development. The ramp-up and maintenance were also taken care of by Nokia for most applications.

The existing cloud-based data analytics platform had already comprehensively implemented a set of services for different areas. Services for data processing, data storing, data analytics, and monitoring were already in place, and the target was to develop a new microservice. This monitoring service was intended to be a new functionality in the monitoring part of the platform. The use of existing solutions was the preliminary step towards the development.

The site reliability team had started to put more focus on predictive maintenance of the platform so there was a need for an application that would monitor and predict the health and behavior of different components of the platform. By using machine learning techniques, the predictability of the platform could be improved thus providing more control over the maintenance the platform.

In addition, significant costs reduction could be achieved by decreasing cost that are caused by unexpected behavior of the platform.

4.2 Application to be integrated

The target application in the integration was intended to be open-source software. Most of the components of the platform were open-source software running inside of OS-based containers. The platform is based on an orchestration engine powered by Apache Marathon and Apache Mesos.

Key elements for the anomaly detection system are data streaming, data processing, and data visualization. The platform already included required components for data streaming and data visualization, which were Apache-powered time-series database for data streaming (Prometheus) and Apache-powered data visualization dashboards (Grafana). The data processing would be the developed application itself, and an addition to the core functionalities of the platform. From this chapter and on the Prometheus services, will be called **Telemetry server** and Grafana **Telemetry Dashboard**.

The major assumption at the beginning of the development process was that there are existing solutions fitting these previously mentioned technologies.

4.3 Planning and development process

The planning and development work was following agile Scrum principles. First, initial requirements were developed in a sprint planning meeting, and tasks were created relating to the work. Generic discussion about the topic was held between the entire team, and because of the spring planning, there were initial tasks generated to the backlog for the microservice to be developed.

The second step of the development was research on the subject matter. In the sprint planning meeting, there was a detailed research task created to investigate anomaly detection existing solutions in the context of cloud-based applications, and especially solutions relating to the Telemetry server. During the research phase, a lot of discussions were held with teams working with the platform, and knowledge exchanged regarding different solutions for anomaly detection predicting. After some research, Prometheus compatible open-source solution was found which

seemed to fit the purpose. The solution was Python programming language-based training pipeline that uses two different prediction models Prophet (python library developed by Facebook Inc.) and Fourier Extrapolation.

The third step of the development was to integrate the found solution to the platform. A lot of testing was done locally before deployment. The training was tested and its prediction accuracy using historical data generated by the Telemetry server. From a technical point of view, different parameters of the model were tested and their impact on the predictions was analyzed.

The last step of the development was to configure and deploy the microservice with the platform. This required a lot of discussions with the team regarding the requirements for the integration. The requirements were identified and configured to the platform and ultimately the application was deployed (explained in more detail in the next chapter). After the integration, the application was tested in a production environment, how well it performs, and model parameters were tuned to meet the relevance of the use case. Finally, the work was documented, and usage instructions were provided in the code repository.

4.4 Deployment view

Like previously mentioned, the application is deployed with **Marathon**/**Mesos** container orchestration solution. The deployment of the application itself requires 3 steps: configuring of environmental variables, deployment on Marathon, and creation of route to visualization.

In the deployment of the application, environment variables must be defined. The required variables will be used to communicate with the **Prometheus** and configuring the training of the models. Some of the variables have default values but should be set to be specified by the developer depending on the use case, and the metric that is being monitored. Table 1 below contains all the environmental variables that should be configured when deploying the application:

Table 2 Configuration parameters of the application

Name	Default	Description
METRIC_NAME	`N/A`	Metric name to be forecasted
LABEL_CONFIG	`N/A`	Label values of a specific metric
PROMETHEUS_URL	`N/A`	Prometheus host URL
CHUNK_SIZE	`1h`	download the complete data in smaller chunks, should be less than or equal to DATA_SIZE
DATA_SIZE_HOUR	`12h`	Net data size to scrape from Prometheus
TRAINING_REPEAT_HOURS	`3h`	Iteration time that the model will be retrained
PREDICTION_RANGE	`180`	Range of predictions from the current time as minutes
HTTPS_PROXY	`http://fihel1d-proxy.emea.nsn-net.net:8080`	Proxy URL

Deployment on **Marathon** is done using **Marathon** UI and configuring JSON template which defines specific parameters relating to the cloud environment, like how much CPU, memory, or disk space the application allocates from the platform. See the template below.

```
{
  "id": "/prometheusad",
  "cmd": null,
  "cpus": 3,
  "mem": 3072,
  "disk": 0,
  "instances": 1,
  "constraints": [
    [
      "hostname",
      "UNIQUE"
    ]
  ],
  "acceptedResourceRoles": [
    "*"
  ],
  "container": {
    "type": "DOCKER",
    "docker": {
      "forcePullImage": true,
      "image": "ava-docker-local.esisoj70.emea.nsn-net.net/ava/monitoring/metric-
based-anomaly-detection:0.1.3",
      "parameters": [
        {
          "key": "log-opt",
          "value": "labels=app"
        },
        {
          "key": "label",
          "value": "app=prometheusAD"
        },
        {
          "key": "label",
          "value": "namespace=ava-system"
        }
      ],
      "privileged": false
    },
    "volumes": [],
    "portMappings": [
      {
        "containerPort": 8080,
        "labels": {
          "metrics": "/metrics"
        },
        "name": "default",
        "protocol": "tcp",
        "servicePort": 10044
      }
    ]
  },
  "healthChecks": [
    {
      "gracePeriodSeconds": 300,
      "intervalSeconds": 60,
```

As soon as Marathon is running the container, the route to Flask server endpoint must be created using **Kong API Gateway**. The **Kong API gateway** service is hosted on the platform and its responsibility is to manage the connectivity between the services on the platform. The **Kong service** is used for the creation of the route between the application and visualization endpoint (Flask server). This is simply done using the following command on the instance host machine:

```
curl -X POST server-kong.marathon.mesos:8001/apis/ -d 'name=prometheusd' -d
'upstream_url=prometheusad.marathon.mesos:8080/plots' -d 'uris=/plots' -d
'strip_uri=true'
```

4.5 Functional view

This application will provide time-series forecasts on Prometheus metrics, and compare the predicted values with actual values. The application will detect an anomaly data and flag it as anomaly if the actual metric values are out of the boundaries of predictions.

The application leverages machine learning algorithms of two types, Fourier Extrapolation and Prophet. The metrics data will be run through both algorithms and predictions are generated for each one independently. The output of the predictions is 6 different metrics, which are then scraped by Prometheus Service Discovery from the Flask web server the application is hosting. This application will provide two graphs of the forecast, on the Flask web server endpoint.

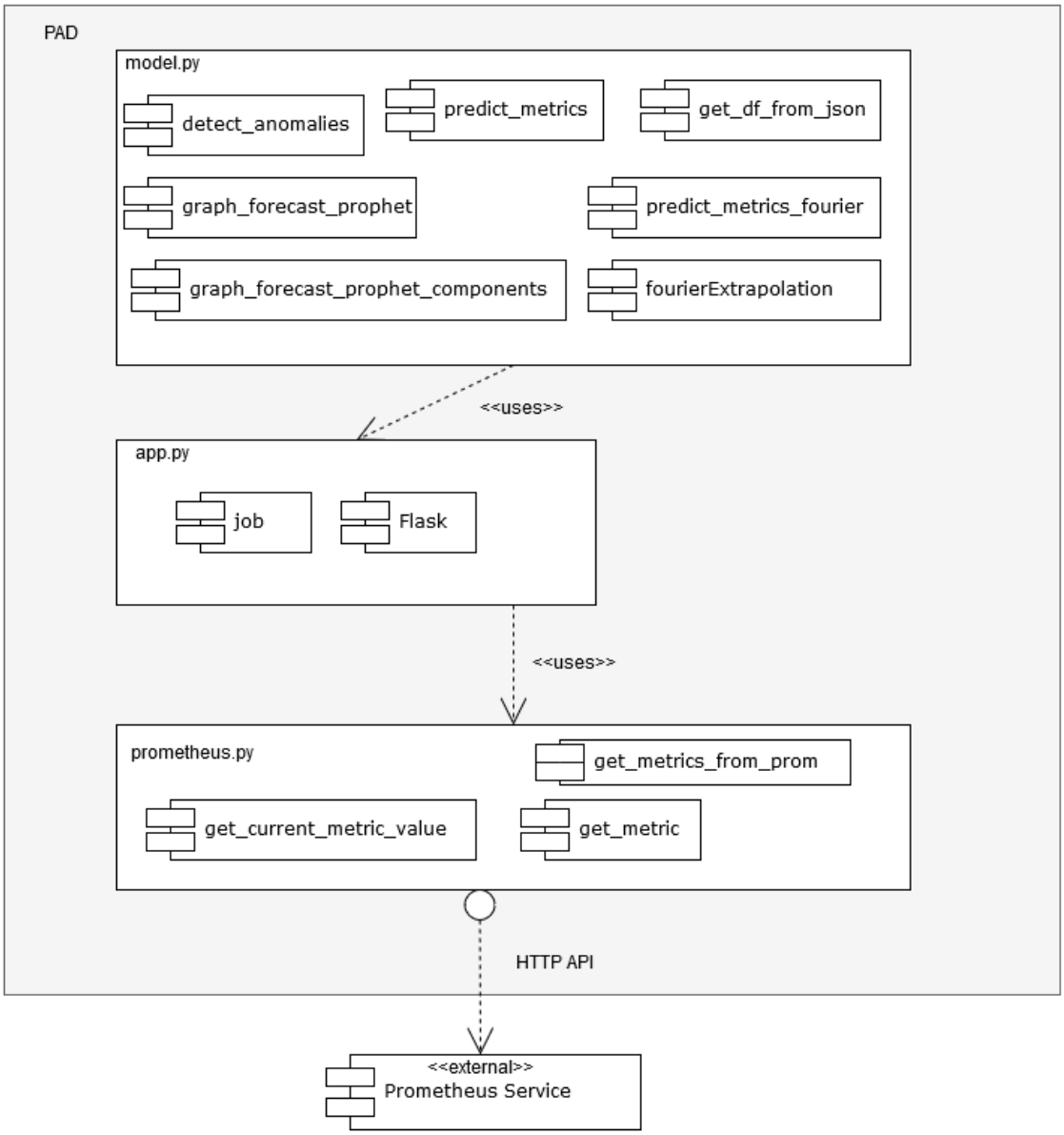


Figure 6 UML diagram of anomaly detection application

When the app is started, on TRAINING_REPEAT_HOURS interval, the model is trained.

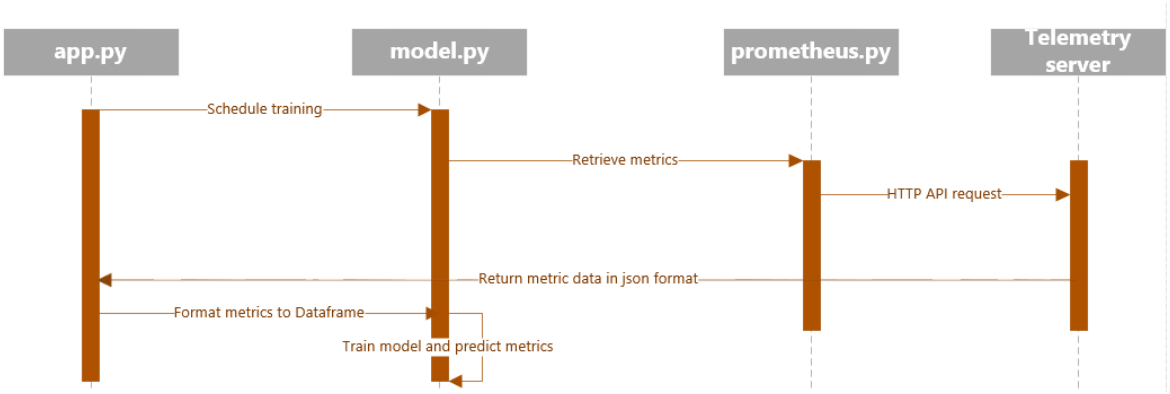
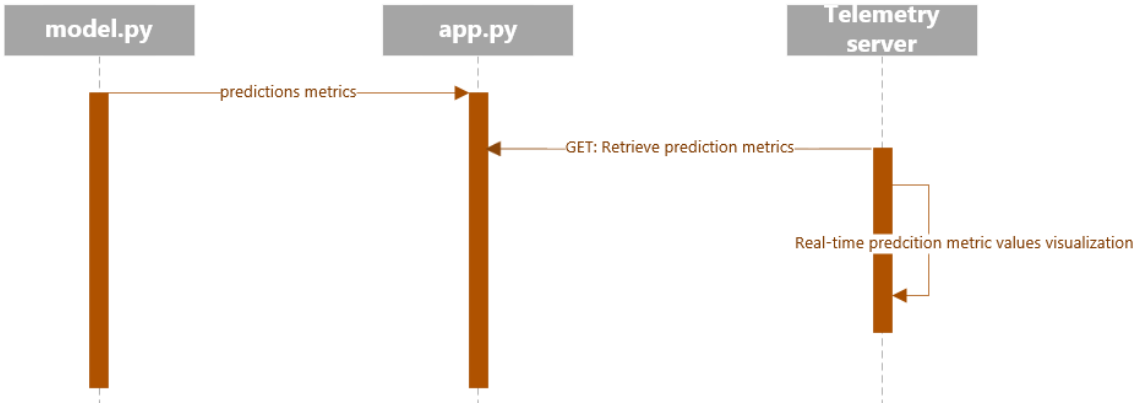


Figure 7 Training the Prophet model

The application exposes /metrics endpoint for Prometheus to scrape the prediction metrics for interactive graphs.



After the application is started and the model is trained the application serves prediction plots as an HTML page in /plots endpoint.

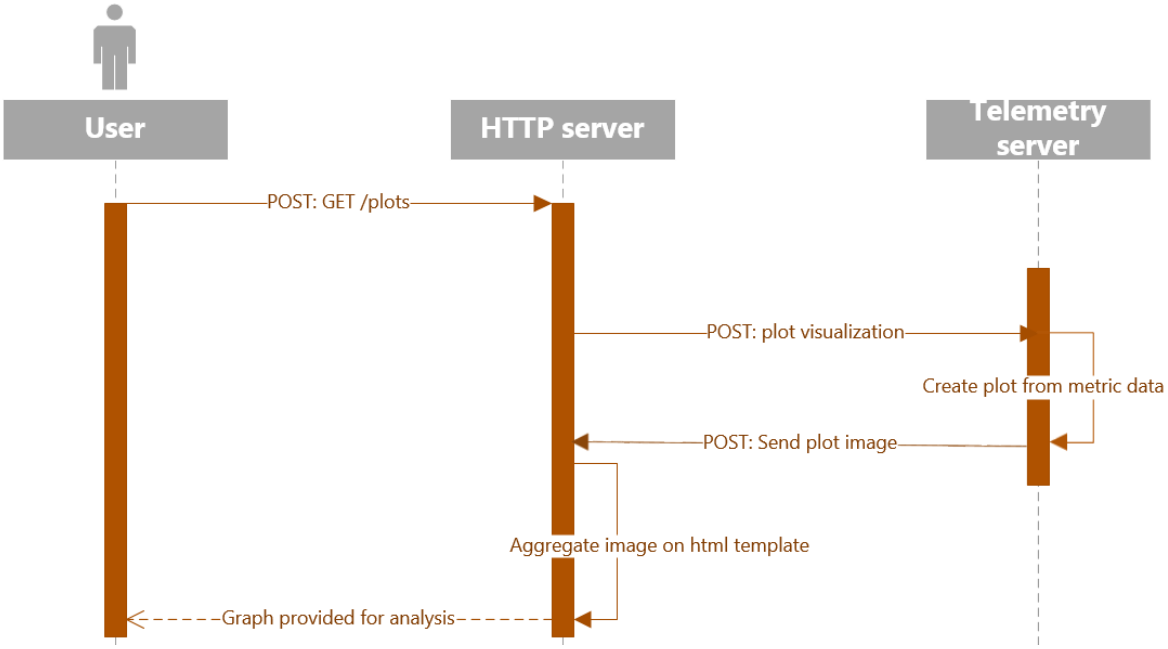


Figure 8 Serving plots of the anomaly detection application.

4.6 Application usage

The use case for this framework is to support the maintenance team in real-time anomaly detection of their system components. The time series forecasts can be used by developers to detect unexpected behavior of metrics. However, the developer should have domain expertise on the monitored metric, to be able to understand the meaning of metric values. Alerting of anomaly data is not automated but powered by visualization. The application can be used for monitoring one instance of a metric at a time. As previously mentioned, the user is required to specify *label_config* of a metric the model will be trained on.

4.6.1 Inspecting logs

After successful deployment with Marathon, the logs will be accessible via Mesos. The log output is presented in the picture below.

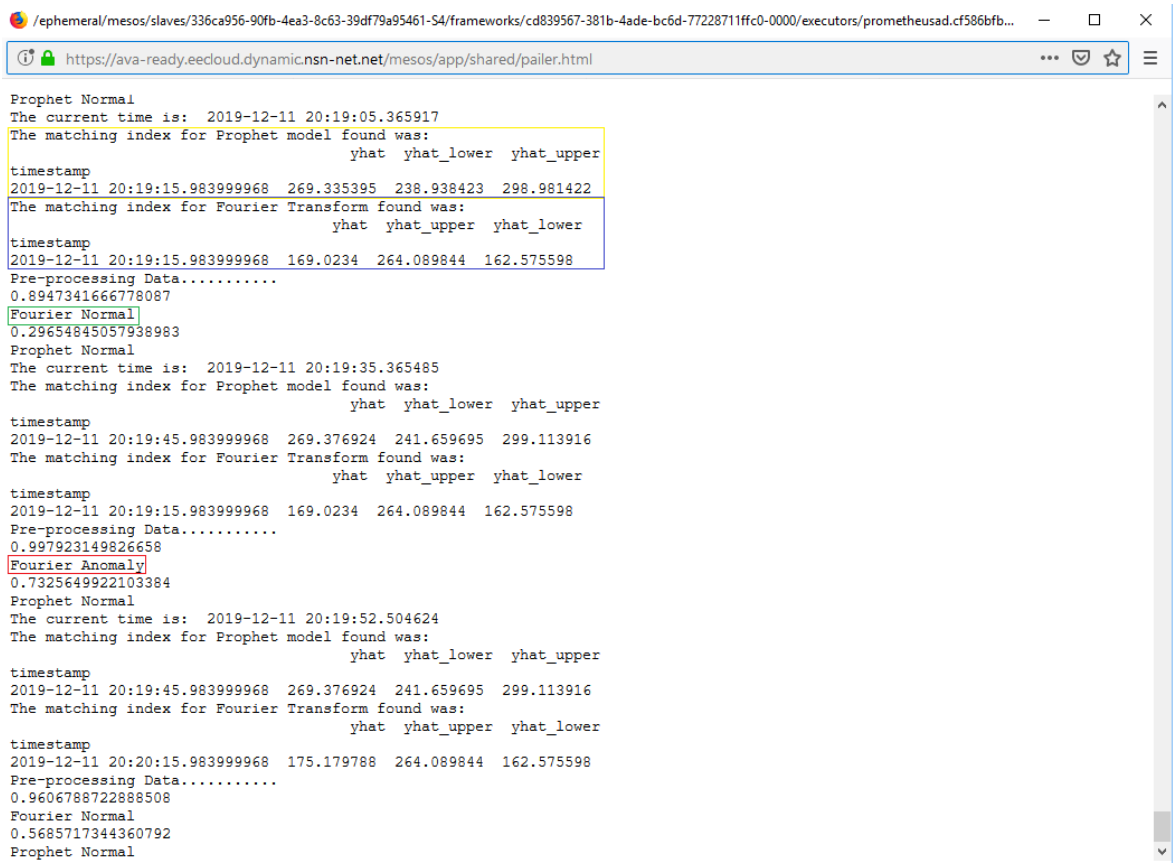


Figure 9 Real-time logging of the application

Inside the yellow box is Prophet predicted value (yhat), prediction upper limit (yhat_upper), and prediction lower limit (yhat_lower). Inside the blue box is Fourier predicted value (yhat), prediction upper limit (yhat_upper), and prediction lower limit (yhat_lower). If the current metric value is what is expected, the application will print the text as shown inside the green box. If the current metric value is considered an anomaly, the application will print the text as shown inside the red box.

After the model has been trained, the application will also expose the predictions as new metric data for Prometheus to scrape through Prometheus Service Discovery automatically. The user can access the real-time prediction metrics from the Telemetry server. Telemetry server UI in the picture below.

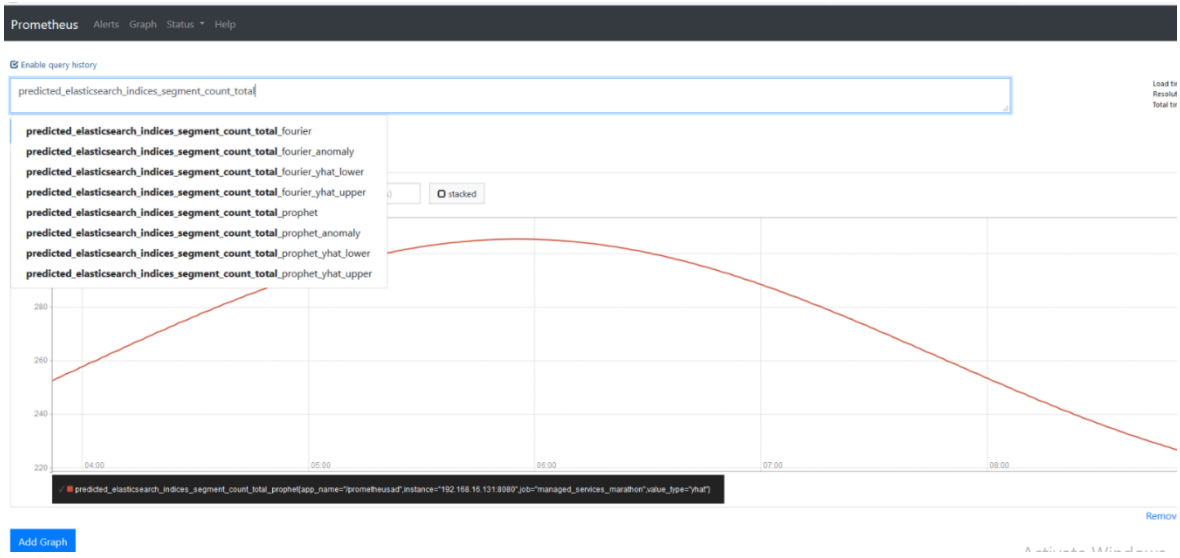


Figure 10 Telemetry server UI (Prometheus)

4.6.2 Visualization of predictions

When the application is deployed on Marathon, to display visualization, a route to plots endpoint must be created manually. After the route is created to *PAD_ENDPOINT*, the model plots should be available at HTTPS://<PaaSHWAddress>/plots. The visualization will provide 2 different graphs generated from the model: forecast graph with actual metric values with upper and lower, and mean values (Figure 10), and forecast seasonality (Figure 11).

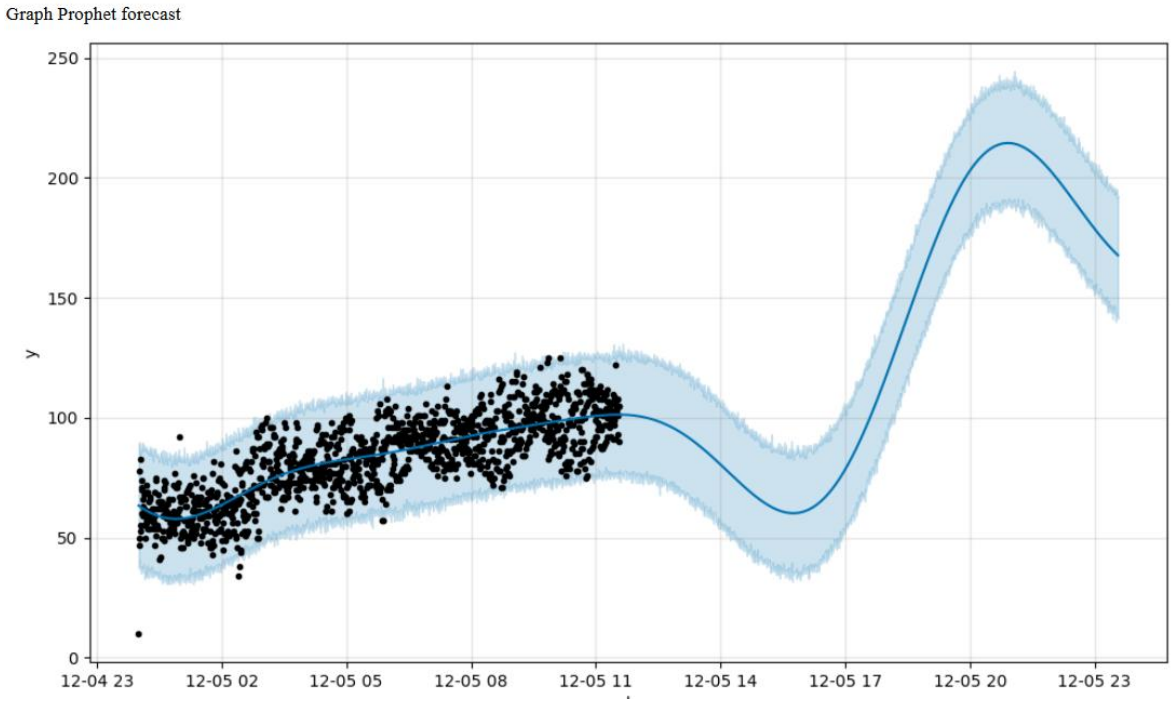


Figure 11 Actual values within the prediction boundaries

In Figure 10, the black dots are actual metric values, the blue line being the median values, and the light blue area being the expected boundaries where the actual values should be positioned.

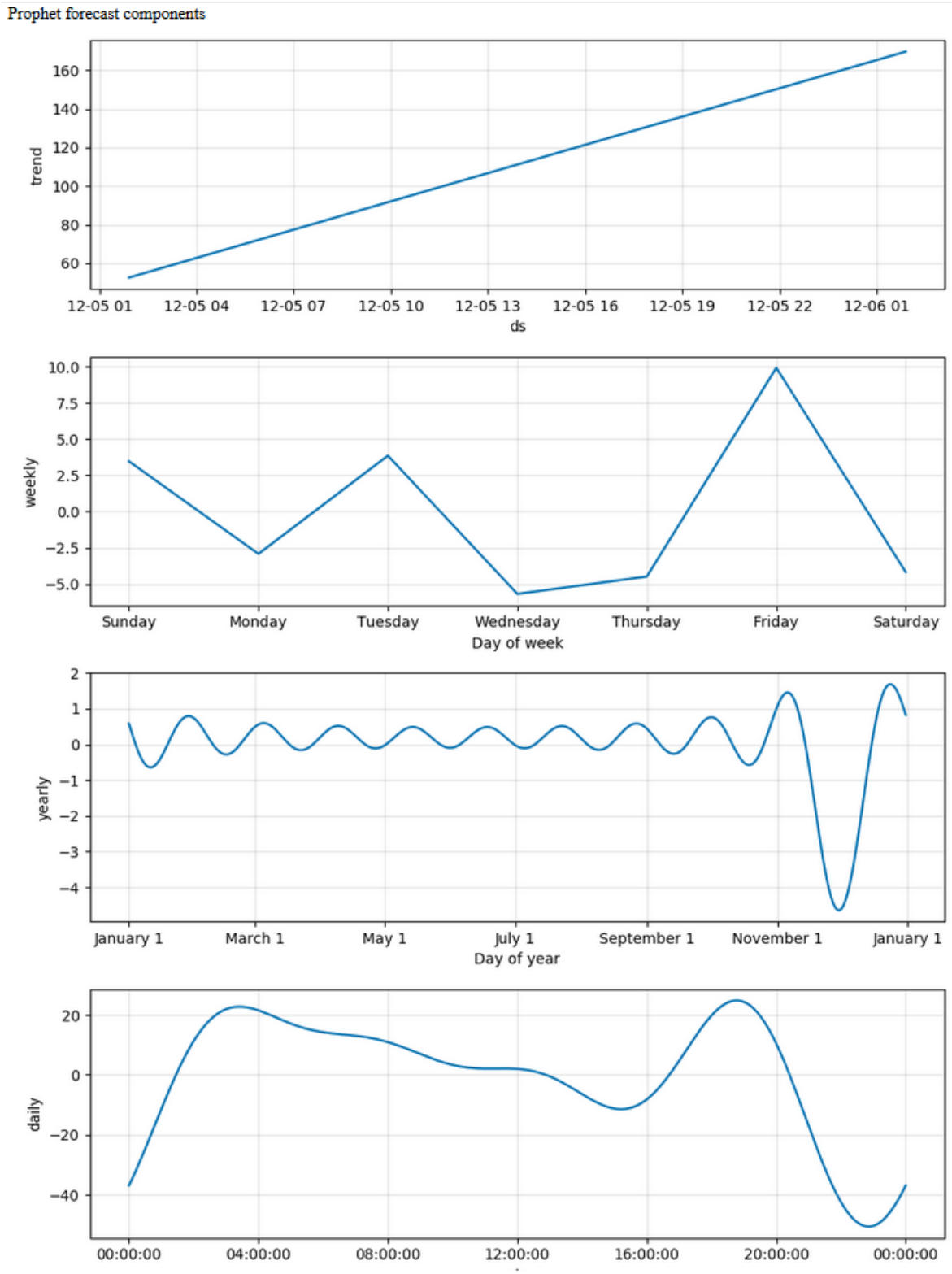


Figure 12 Forecast as seasonality components

4.7 Future development proposals

The AVA project was ramped down the end of December 2019, at the time of acceptance testing of the application, and the development was left a bit incomplete. For example, there was no time to create Telemetry Dashboard (Grafana) based visualization for the predictions, instead, the visualization was implemented using a separate HTTP server that would expose them as an image. As a future development proposal, integration to Telemetry Dashboard would be worthy attention.

Also, the data processing of the application was initially designed to be real-time data streaming, however, the integrated solution was leveraging batch processing (Python-based asynchronous processing) of data which decreases the availability and accuracy of the predictions. As a development proposal, there could be added some new data processing feature or component that makes the training pipeline real-time data stream based such as Apache Flume agents.

However, the AVA project was ramped down at the end of the development work, so further development ideas, in this case, would not make sense.

5. DISCUSSION AND CONCLUSIONS

Some of the challenges found in the literature review were faced during the development work, such as data visualization, nature of the input, parameters selection, and accuracy. In fact, data visualization (the adoption of **Telemetry Dashboard**) turned out to be more challenging than expected during the planning. Embedding open-source software seems to have limitations with compatibility issues, and in some cases, it turns into a more complex problem as there emerges a need to develop new functionality to integrate some tools together. Concerning the implementation, because of the limited time, the use of the **Telemetry Dashboard** was intentionally dropped off from the backlog so that minimum viable product could be ensured to be delivered on time. As a future development idea, the integration to **Telemetry Dashboard** would be the next thing to consider, to get an interactive visualization for the application.

Another time-consuming challenge during the integration was parameter selection. Optimal parameter values of the model were hard to find, due to the heavy load of regression testing. This would have required more time, planning and a systematic way of testing, as well as proper documentation since most of the testing, was done manually. Alternatively, automated testing could have improved the test coverage, and do the job.

To make the developed application reliable and business-wise sustainable, the application would require much more testing than had been done during the work. Also, the model is not a fixed version that could be always used – instead, it requires systematic and active maintenance activities to keep it relevant – which would lead to more resources allocated to it as well.

Most of the challenges mentioned in the literature were identified during the development work, and so the conducted literature review gave a good pre-caution about what must be prepared for and thus helped to improve the overall performance of the development work.

During the development work, the importance of **active** teamwork and **active** discussions came to light as most of the ideas and technical details were generated from meetings and discussions with

colleagues. The development work also gives also a good overall picture on the way of working in software company. The implementation work was mostly configuration management and testing, and it gives fresh graduates an idea what future work would be like.

Anomaly detection and monitoring, in general, are a complex problem which requires a lot of resources and planning. If one wants to implement monitoring or anomaly detection services, it requires extraordinary talent and a lot of resources.

6. SUMMARY

In this research, the goals were to generate understanding in the fields of predictive maintenance and anomaly detection and implement anomaly detection application with AVA cognitive services platform. A literature review was conducted to cover the background knowledge on PdM and anomaly detection and the related challenges and techniques. The practical work was presented through the AVA integration experience report.

PdM and anomaly detection are complex problems and require a lot of technical expertise in areas of data processing, data storing and data streaming. The main challenges of anomaly detection are technical details such as the nature of data, the accuracy of models, and other previously mentioned details. However, studies show that organizations are working for tools and techniques to make anomaly detection monitoring services easier to implement in the future.

The literature reveals that PdM and anomaly detection has the potential to improve machinery lifetime and maintainability by providing a more interactive way of working. It can also reduce operational costs, improve flexibility and performance.

From the development process point of view, the key to a successful PdM strategy is active and engaging communication skills. The requirements management of such services is demanding and there are a lot of hidden requirements which require the use of some standardized development process model and determination to execute it.

Future development suggestions regarding the application include integration to Grafana dashboards and more testing and optimization of the models.

REFERENCES

1. Gartner. 2019. Gartner Forecasts Worldwide Public Cloud Revenue to Grow 17% in 2020.
2. Durcevic, S. 2020. The Top 10 SaaS Trends for 2020 That Will Disrupt the Industry.
3. Enterprisersproject.com. 2020. What is digital transformation?.
4. Wang, K., 2020. Intelligent Predictive Maintenance (Ipdm) System – Industry 4.0 Scenario.
5. Tran Anh, D., Dąbrowski, K., Skrzypek, K. 2018. The Predictive Maintenance Concept in the Maintenance Department of the "Industry 4.0" Production Enterprise, Foundations of Management, 10(1), pp. 283-292. doi: https://doi.org/10.2478/fman-2018-0022
6. Wang, W., 2020. AVA Cognitive Services Platform | Nokia. [online] Nokia.
7. Lughofer E., Sayed-Mouchaweh M. Predictive maintenance in dynamic systems: Advanced Methods, Decision Support Tools and Real-World Applications (1st ed.). Springer, New York. ISBN 978-3-030-05644-5.
8. Www1.eere.energy.gov. 2020. Operations & Maintenance Best Practices Guide: Release 3.0.
9. UpKeep Inc. 2020. Preventive Vs. Predictive Maintenance: What'S The Difference?
10. Eagle CMMS. 2019. Importance And Benefits Of Predictive And Preventive Maintenance.
11. Parida, V., Sjödin, D., Reim, W., 2020. Reviewing Literature On Digitalization, Businessmodel Innovation, And Sustainable Industry:Past Achievements And Future Promises.
12. Vithayathil, J., 2020. Will Cloud Computing Make The Informationtechnology (IT) Department Obsolete?.
13. Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., d Barbosa, J., 2021. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges.
14. Zhou, C., Tham, C.-K., 2018. Graphel: a graph-based ensemble learning method for distributed diagnostics and prognostics in the industrial internet of things. In: 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), IEEE, pp. 903–909.

15. Liu, Z., Jin, C., Jin,W., Lee, J., Zhang, Z., Peng, C.,Xu, G., 2018. Industrial ai enabled prognostics for high-speed railway systems. In: 2018 IEEE International Conference on Prognostics and Health Management (ICPHM), IEEE, pp. 1–8.
16. Crespo Márquez, A., de la Fuente Carmona, A., Antomarioni, S., 2019. A process to implement an artificial neural network and association rules techniques to improve asset performance and energy efficiency. Energies 12, 3454.
17. Hegedüs, C., Varga, P., Moldován, I., 2018. The mantis architecture for proactive maintenance. In: 2018 5th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE, pp. 719–724.
18. Cerquitelli, T., Bowden, D., Marguglio, A., Morabito, L., Napione, C., Panicucci, S., Nikolakis, N., Makris, S., Coppo, G., Andolina, S., et al., 2019. A fog computing approach for predictive maintenance. In: International Conference on Advanced Information Systems Engineering, Springer, pp. 139–147.
19. Calabrese, M., Cimmino, M., Fiume, F., Manfrin, M., Romeo, L., Ceccacci, S., Paolanti, M., Toscano, G., Ciandrini, G., Carrotta, A., et al., 2020. Sophia: an event-based iot and machine learning architecture for predictive maintenance in industry 4.0. Information 11, pp. 202.
20. Strauß, P., Schmitz, M., Wöstmann, R., Deuse, J., 2018. Enabling of predictive maintenance in the brownfield through low-cost sensors, an iiot-architecture and machine learning. In: 2018 IEEE International Conference on Big Data (Big Data), IEEE, pp. 1474–1483.
21. Schmidt, B., Wang, L., 2018. Predictive maintenance of machine tool linear axes: a case from manufacturing industry. Proc. Manuf. 17, pp. 118–125.
22. ADHIKARI, P., RAO, H.G., BUDERATH, D.-I.M., 2018. Machine Learning Based Data Driven Diagnostics & Prognostics Framework for Aircraft Predictive Maintenance.
23. Schmidt, B., Wang, L., 2018. Cloud-enhanced predictive maintenance. Int. J. Adv. Manuf. Technol. 99, pp. 5–13.
24. Yuan, Y., Ma, G., Cheng, C., Zhou, B., Zhao, H., Zhang, H.-T., Ding, H., 2018. Artificial Intelligent Diagnosis and Monitoring in Manufacturing. arXiv preprint arXiv:1901.02057
25. Peres, R.S., Rocha, A.D., Leitao, P., Barata, J., 2018. Idarts-towards intelligent data analysis and real-time supervision for industry 4.0. Comput. Ind. 101, 138–146.

26. Li, Z., Wang, Y., Wang, K.-S., 2017. Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario. Adv. Manuf. 5, pp. 377–387.
27. Nunez, ˜ D.L., Borsato, M., 2018. Ontoprog: an ontology-based model for implementing prognostics health management in mechanical machines. Adv. Eng. Inform. 38, pp. 746–759.
28. Q. Cao, A. Samet, C. Zanni-Merk, F. d. B. de Beuvron, C. Reich, 2020. Combining chronicle mining and semantics for predictive maintenance in manufacturing processes. 11, pp. 927-948.
29. Hegedüs, C., Varga, P., Moldován, I., 2018. The mantis architecture for proactive maintenance. In: 2018 5th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE, pp. 719–724
30. Nokia. 2020. Nokia AVA Telco AI Ecosystem | Nokia. [online] Available at: <https://www.nokia.com/networks/solutions/nokia-ava-telco-ai-ecosystem/>
31. Nokia. 2020. Nokia Launches AVA 5G Cognitive Operations To Help Telcos Enter The 5G Era | Nokia. [online] Available at: <https://www.nokia.com/about-us/news/releases/2020/03/31/nokia-launches-ava-5g-cognitive-operations-to-help-telcos-enter-the-5g-era/>
32. Nokia. 2020. Nokia AVA Analytics Services | Nokia. [online] Available at: <https://www.nokia.com/networks/services/nokia-ava-analytics-services/#nokia-ava-platform>
33. Nokia. 2020. Nokia AVA, The New Cloud-Based Cognitive Platform For Fast, Flawless Service Delivery To Operators #MWC16 | Nokia. [online] Available at: <https://www.nokia.com/about-us/news/releases/2016/02/08/nokia-ava-the-new-cloud-based-cognitive-platform-for-fast-flawless-service-delivery-to-operators-mwc16/#:~:text=Nokia%20AVA%20is%20a%20cloud,faults%20and%20solves%20them%20rapidly.>
34. Nokia. 2020. Analytics And AI Services | Nokia. [online] Available at: <https://www.nokia.com/networks/services/analytics-and-ai-services/>
35. Habeeb, R.A., Nasaruddin, F., Gani, A., Hashem, I.A., Ahmed, E., & Imran, M. 2019. Real-time big data processing for anomaly detection: A Survey. Int. J. Inf. Manag., 45, 289-307.

36. Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., Amira, A. 2021. Artificial intelligence-based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives.
37. Chandola, V., Banerjee, A., & Kumar, V. 2009. Anomaly Detection: A Survey. ACM Computing Surveys.
38. Adamova, K., Schatzmann, D., Plattner, B. 2012. Network Anomaly Detection in the Cloud: The Challenges of Virtual Service Migration. Austrian Institute of Technology.
39. Bose, S.K., Kar, B., Roy, M., Gopalakrishnan, P.K., Basu, A. 2019. ADEPOS: anomaly detection-based power saving for predictive maintenance using edge computing. Proceedings of the 24th Asia and South Pacific Design Automation Conference.
40. Chuadhry, A., Gauthama, R., Aditya, M., 2021. Challenges in Machine Learning based approaches for Real-Time Anomaly Detection in Industrial Control Systems | Proceedings of the 6th ACM on Cyber-Physical System Security Workshop.