WP4 – DIGITALIZATION OF OPERATIONAL PROCESSES, SMART SENSORS CONFIGURATION AND DSS IMPLEMENTATION

Task 4.3 – DSS Analytics Services

Person responsible / Author:	Dimitris Ntalaperas (SINGULAR)
Deliverable No.:	4.3
Work Package No.:	WP4
Date:	20.12.2019
Project No.:	760792
Classification:	Public

File name:	FENIX_D4.3 report_v1
Number of pages:	17

Status of deliverable

Action	By	Date (dd.mm.yyyy)
Submitted (author(s))	Dimitris Ntalaperas (SINGULAR)	20.12.2019
Responsible (WP Leader)	Dimitris Ntalaperas (SINGULAR)	

Revision History

Date (dd.mm.yyyy)	Revision version	Author	Comments
28.11.2019	v0.1	Dimitris Ntalaperas, (SINGULAR)	ToC Definition
13.12.2019	v0.5	Dimitris Ntalaperas, (SINGULAR)	
20.12.2019	v0.8	Dimitris Ntalaperas (SINGULAR)	Draft version
31.12.2019	v0.8	Claudio Sassanelli, Paolo Rossa (POLIMI) Fransesco Veglio (UNIVAQ)	Review
13.01.2020	v1.0	Dimitris Ntalaperas (SINGULAR)	Final Version

Author(s) contact information

Name	Organisation	E-mail	Tel
Dimitris Ntalaperas	SINGULAR	dimitris.ntalaperas@gmail.com	+30 6944029042

ABSTRACT

The current document reports the development of the analytics services which can be used both as standalone tools to provide insights upon deeper characteristics of the collected data as well as supporting modules to the DSS Engine to generate better suggestions. Although the analytics services are to be used with data that will be collected during the pilot demonstrators that will take place in WP6, the main requirements have already been collected in collaboration with the pilots; these have been crystallized into a set of core services that would provide the best support digitizing the recycling process. These services are time series forecast, outlier detection and pattern recognition, and are being documented in the present deliverable.

Table of Contents

Figures

Tables

Abbreviations and Acronyms:	
CBM	Circular Business Model
DSS	Decision Support System

INTRODUCTION

Deliverable 4.3: DSS Analytics Services documents the specification and implementation of the Analytics services that are to be used in the context of FENIX to extract insight from the characteristics of the data collected, both streaming and batch, during the pilot execution of FENIX. The structure of the present document can be summarized as follows:

- Section 1 builds upon the requirement engineering process defined in D4.2; it gives a summary of the updates to the requirements as these were defined during the activities of T4.3
- Section 2 refines the conceptual architecture that was defined for the whole DSS in D4.2. The refinements needed for the Analytics Services are documented in the present section, while the refinements needed for the DSS Engine are documented in D4.4
- Section 3 documents the services that were implemented and the main outcomes of the present work
- Section 4 gives a brief outline on how the Analytics Services can be used and integrated to the DSS platform
- Section 5 finally, gives the main conclusions of the present work.

1. REQUIREMENTS ENGINEERING

The current section gives an overview of the business requirements of WP1, shows how they are mapped to technical requirements and, how the last are reflected to the current architecture of the DSS Analytics Services. This work was documented in detail in D4.2; in the present document the requirements applicable to the DSS Analytics Services are analysed.

1.1. Functional Requirements

Table 1 presents the core functional requirements of the FENIX DSS Platform, as they are derived from the analysis of the WP1 outcomes, as well as, T4.1. The unique ID of the requirement has been provided for each functional requirement, a brief description of what this requirement entails, the component that should address the requirement and, finally, the source from which this requirement originates (a use case or the results of a Task). The requirements that are relevant for the DSS Analytics Services, have been highlighted in green colour.

In essence, the main role of the Analytics Services is to try and identify and compute metrics and trends that are representative of the recycling processes. Using these estimations (both current values and projections), will help with the generation of optimized suggestions regarding operator actions.

Table 1: Functional Requirements

ID	Need	Description	Component	Input/Use case
FR1	Stream data processing	FENIX DSS Platform should be able to process stream data	DSS Data Fusion	D1.2

FR2	Legacy data processing	FENIX DSS Platform should be able to process legacy data	DSS Data Fusion	D1.2
FR3	Processing data from web sources	FENIX DSS Platform should be able to process data from web sources (social web platforms)	DSS Data Fusion	UC2
FR4	Energy data processing	FENIX DSS Platform should be able to define, access and process energy data	DSS Data Fusion	D1.2
FR5	Thermal data processing	FENIX DSS Platform should be able to define, access and process thermal data	DSS Data Fusion	D1.2
FR6	Materials data processing	FENIX DSS Platform should be able to define, access and process material data	DSS Data Fusion	D1.2, UC3
FR7	Resources data processing	FENIX DSS Platform should be able to define, access and process resources data	DSS Data Fusion	D1.2
FR8	Estimation of energy consumption	FENIX DSS Platform should assist in the estimation of energy consumption	DSS Data Fusion, DSS Engine	D1.2
FR9	Estimation of recyclability index	FENIX DSS Platform should assist in the estimation of the material recyclability	DSS Analytics, DSS Engine	D1.2 (recyclability index)
FR10	Estimation of circularity indicators	FENIX DSS Platform should assist in the estimation of the circularity indicators	DSS Analytics, DSS Engine	D1.2 (circularity indicators)
FR11	Product Requirements definition	FENIX DSS Platform should provide the ability to define the product requirements from customer's side	DSS Engine	UC1
FR12	Correlation models definition	FENIX DSS Platform should provide the ability to define rules and correlations between the recycled raw material characteristics to final dowser properties	DSS Engine	UC1
FR13	Real time user validation support	FENIX DSS Platform should provide the ability to define rules based on user validation and assessment input of the business models	DSS Engine	UC3
FR14	Health and safety assessment	FENIX DSS Platform should provide the ability to support decision support for health and safety assessment	DSS Analytics	UC3
FR15	Provision of performance-based models	FENIX DSS Platform should provide performance-based models and	DSS Analytics	WP1, all use cases

	and metrics about regarding efficiency and sustainability	metrics about regarding efficiency and sustainability		
FR16	Provision of patterns recognition techniques	FENIX DSS Platform should provide patterns recognition techniques for recognizing specific trends	DSS Analytics	WP1, all use cases
FR17	Provision of algorithms for statistical analysis, trends and forecasting analysis, classification algorithms etc.	FENIX DSS Platform should provide algorithms for statistical analysis, trends and forecasting analysis, classification algorithms etc.	DSS Analytics	WP1, all use cases
FR18	Alerting and signaling mechanisms for patterns analysis and outliers' identification	FENIX DSS Platform should provide alerting and signaling mechanisms for patterns analysis and outliers identification	DSS Analytics, DSS Engine	WP1, all use cases

1.2. Non-Functional Requirements

The process of the identification of non-functional requirement analysis was carried out in D4.2. For completeness, we report here the list of these requirements that will also be used for the identification of the Analytics Services technical requirements. The list can be seen in Table 2

Table 2: Non-functional Requirements (NFR)

ID	Requirement Sub-category	Description
NFR1	Functional Suitability	FENIX DSS Platform should be able to perform a great variety of analysis on selected multiple datasets (private or public) in order to provide state of the art analytics
NFR2	Performance efficiency	FENIX DSS Platform should be able to perform analytics in a timely and efficient manner
NFR3	Performance efficiency	FENIX DSS Platform should be able to guarantee the full optimization of the response time to ensure a functional and flexible navigation
NFR4	Compatibility	FENIX DSS Platform should be able to interact and exchange information with other products / IoT devices in a secure way (for example secure REST API)

NFR5	Usability	FENIX DSS Platform shall feature a user-friendly interface, provide an overview of supported kind of analytics and visualization
NFR6	Reliability	FENIX DSS Platform should ensure high availability of the system and the stored datasets
NFR7	Security	FENIX DSS Platform should offer login with user credentials
NFR8	Portability	FENIX DSS Platform should be able to be deployed in a timely and efficient manner

1.3. Technical Requirements

The technical requirements were driven from the list of functional and non-functional requirements as defined in the previous section and with the mission to utilize the goals and expectations of the users and the stakeholders.

The following list depicted in Table 3 defines the technical requirements that will pave the way for the design and the implementation of Analytics Services. Each technical requirement is mapped to the relevant functional and non-functional requirements, as well as, to one or more architectural sub-components of the Analytics Services to assure that all technical requirements are covered by the component architecture (Figure 1).

Table 3: Technical Requirements

ID	Functional, Non-functional	Need	Description	Sub-component	Users
TR1	FR14, FR18, NFR1, NFR2, NFR6	Health Monitoring	FENIX DSS Platform should be able to detect anomalies in the state of the system	Outlier Detection	D1.2
TR2	FR9, FR10, FR15, NFR1, NFR2, NFR6	Estimation of KPIs	FENIX DSS Platform should be able to provide KPIs to the end users.	All	D1.2
TR3	FR16, FR17, NFR1, NFR2, NFR6	Data Processing	FENIX DSS Platform should be able to process data using arbitrary algorithms	All	D1.2

2. CONCEPTUAL ARCHITECTURE OF THE FENIX DSS PLATFORM

Figure 1 depicts the Conceptual Architecture of the DSS Platform. In summary, the architecture consists of four tiers; the persistency layers for storage, the middle tier that provides interfaces and core services between storage and higher-level functionalities, the enterprise tier which implements

and exposes the main services implementing core business logic and, finally, the presentation tier that integrates all services and provides graphic elements for end users.

For details regarding the functionality of each tier please refer to D4.2

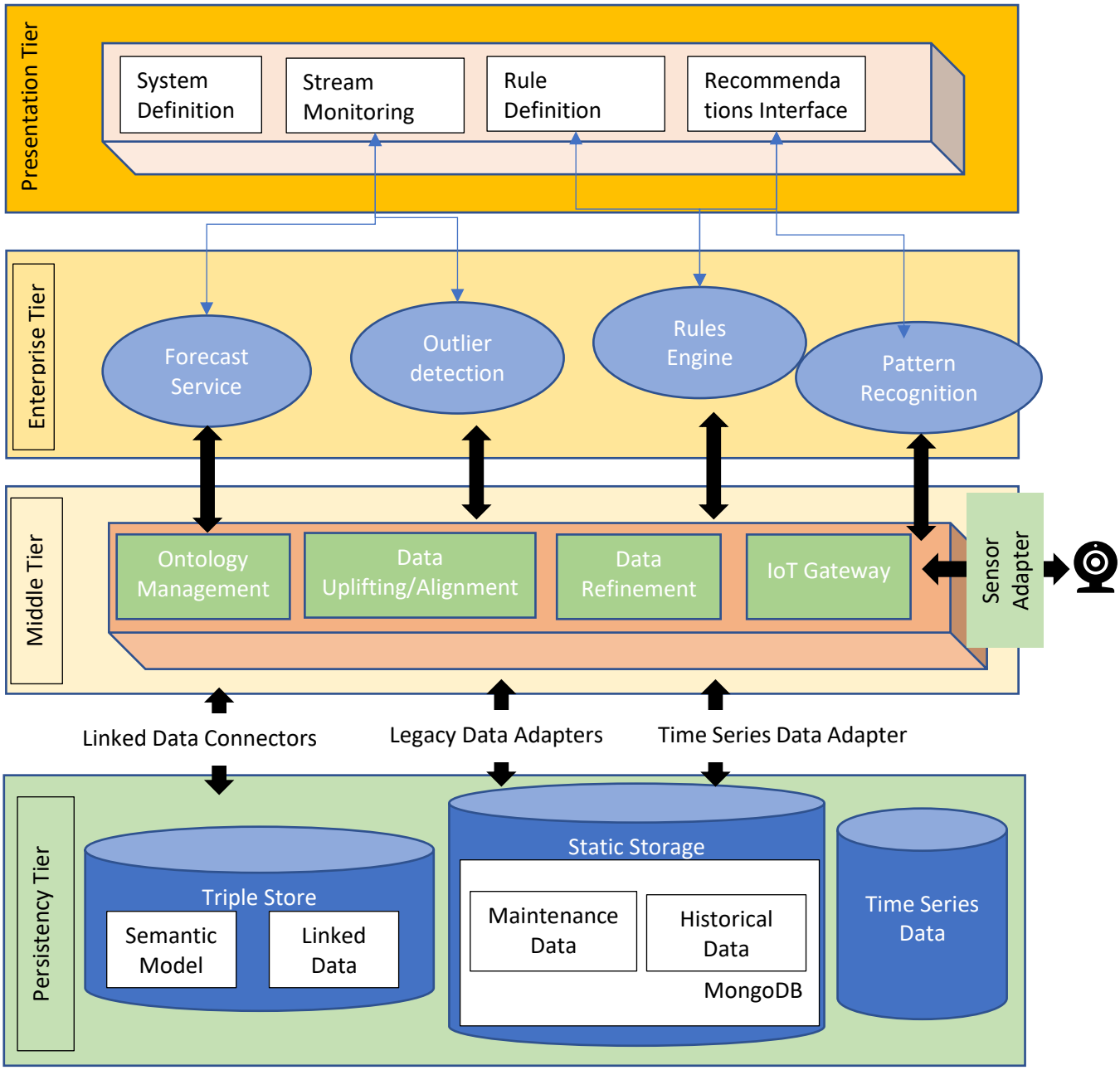

Figure 1: Conceptual Architecture

The architecture has been refined so that the modules of "Alarm Signalling" and "Event handling" have been merged to the single module of "Outlier detection". This occurred due to the fact that it was observed that interesting events (which could lead to the generation of alarms), depended practically upon abnormal behaviour of data. To this end, the functionality of outlier detection has been implemented as a standalone module, with configuration that can lead to specific alarm

generation being delegated to the activities of WP6. This will allow the definition of more precise thresholds that will avoid the generation of false alarm or the suppression of valid ones.

3. ANALYTICS SERVICES MODULES

The Analytics Services consist of the following modules

- Forecast: The forecast module takes as input a series of datapoints in the form of time series and projects as output a series of datapoints that provide a prediction of the future values of the time series
- Outlier detection: This module takes as input a set of datapoints and provides as output a set of outliers datapoints may indicate abnormal behaviour. Outlier detection can ran both on observed data or data that are being predicted by the forecast module, thus providing future alerts
- Pattern recognition: Pattern recognition identifies patterns on the data. This is a functionality that may provide patterns of repetitive behaviour that may give insight to the pilots regarding their recycling capabilities. It is the module most dependent upon the actual data that the pilots will produce during the activities of WP6. It is thus a templated module that will be parametrized and fully tested during the activities of WP6.

In the following subsections we provide a more detailed description of each module

3.1. Forecast

The problem of forecasting time series data is a well-studied one with various techniques being proposed for various domains. It is generally the case that time series data can be forecast with good accuracy for data sources having generic characteristics; even pseudo-random data can be predicted with some accuracy with specific models depending on the existence of some global characteristics of the time series data, such as trend and seasonal components. In general, the accuracy of these methods generally depends both on the nature of data and the size of dataset that is used for training set (please refer to the work cited of Shumway and Stoffer[1] for an overview of the subject of time series analysis).

For FENIX, the exact nature of data produced will be determined during the execution of the pilots. To this end, the forecast module is implemented as a suite of algorithms; the exact algorithm to be executed is defined in the module's configuration. Among the algorithms that are implemented by the module are:

- Autoregression
- Moving Average
- Autoregressive Moving Average
- Exponential Smoothing
- Long Short-Term Memory (LSTM) neural networks

These algorithms were chosen so that time series exposing different characteristics can be handled. Since the data that will be produced by FENIX are not known beforehand, they me exhibit various trends and/or seasonal components. Depending on preliminary analysis that will be performed as soon as the data are produced, these characteristics will be discerned and the most appropriate algorithm will be deployed.

The module is implemented in Python and, as mentioned, uses the template pattern that allows the specification of the exact algorithm at runtime. Figure 2 depicts an example of a forecast using LSTM. For this example, a dataset of the publicly available dataset of OROS has been used.

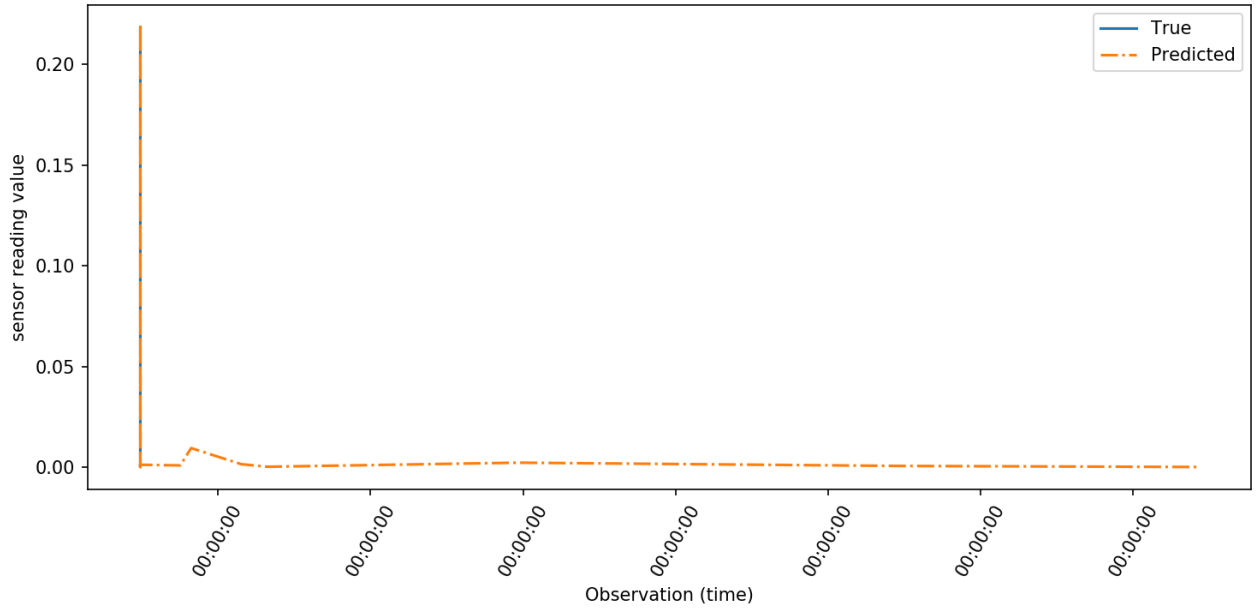

Figure 2: Forecast Example using LSTM

3.2. Outlier Detection

Outliers can indicate points of interest in a dataset that may hint to abnormal behaviour that needs to be addressed. Outliers may not correspond to critical events or catastrophic failures necessarily but may indicate into aspects of the recycling processes that may be improved. For example, spikes in the energy consumption may be due to improper handling of equipment or due to physical characteristics of the specific material being processed. In any case, this outlier can generate an alert to the operator who will evaluate this result.

Outlier detection will be used in tandem with forecasting to produce alerts of future events. Due to the fact that this detection may require a bigger amount of data than that produced in the course of FENIX, it may be the case that it will not be fully validated. However, the methodology will be implemented in a way that can be configurable by the end users so that the appropriate thresholds can be continuously refined.

To implement the prototype of the Outlier Detection module and validate it as a Proof of Concept, a set of pre-generated sensor data was used that was retrieved from OROS. To provide a better demonstrator for the validity of the approach the data were manipulated so that various zones exhibit values away from the average. In particular, the data of sunspot occurrence of the previous three centuries was used as a seed to transform the data. Though the resulting dataset is not realistic anymore, in terms of representing vibration sensor values, it is a dataset that can be efficiently used for training and validating the approach. Figure 3 shows some sample data that were used for implementing the outlier detection algorithm.

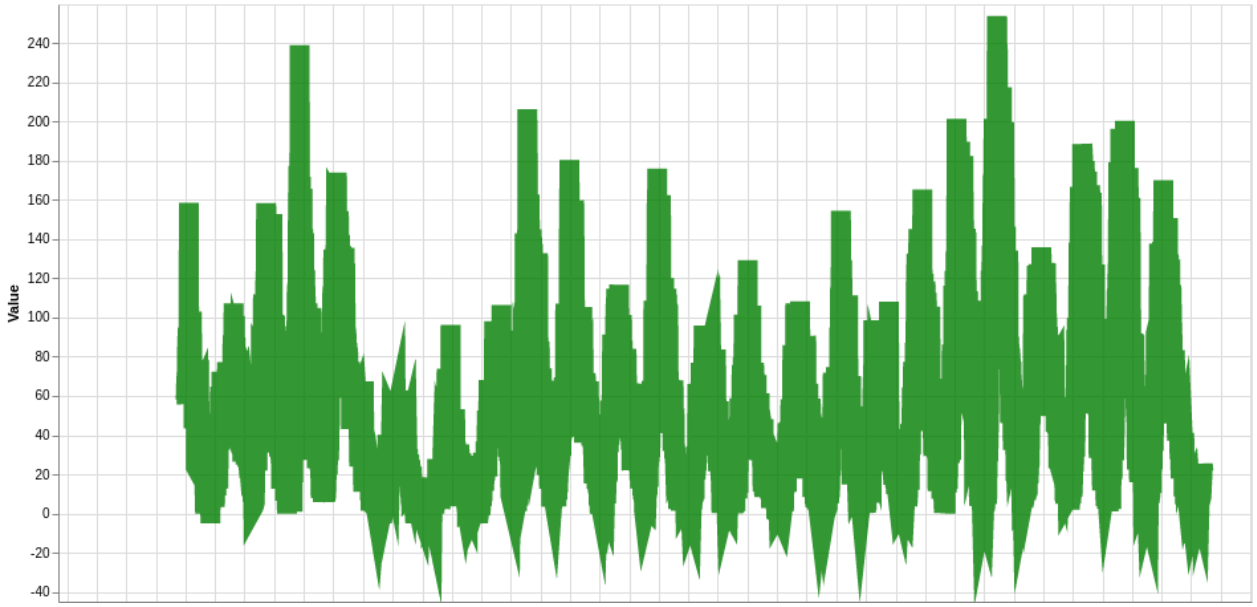

Figure 3: Vibration data sets manipulated to exhibit outliers. Averages of values were taken between days

There are various methodologies for detecting outliers. Since the definition of anomalous behaviour will depend on the data that will be produced in WP6, we do not have enough information to identify anomalous data points beforehand and, therefore, the problem is treated as an unsupervised problem. The module can be configured to use a variety of techniques for outlier detection, such as Isolation Forest or LSTM. Figure 4 depicts the output from the sample data set.

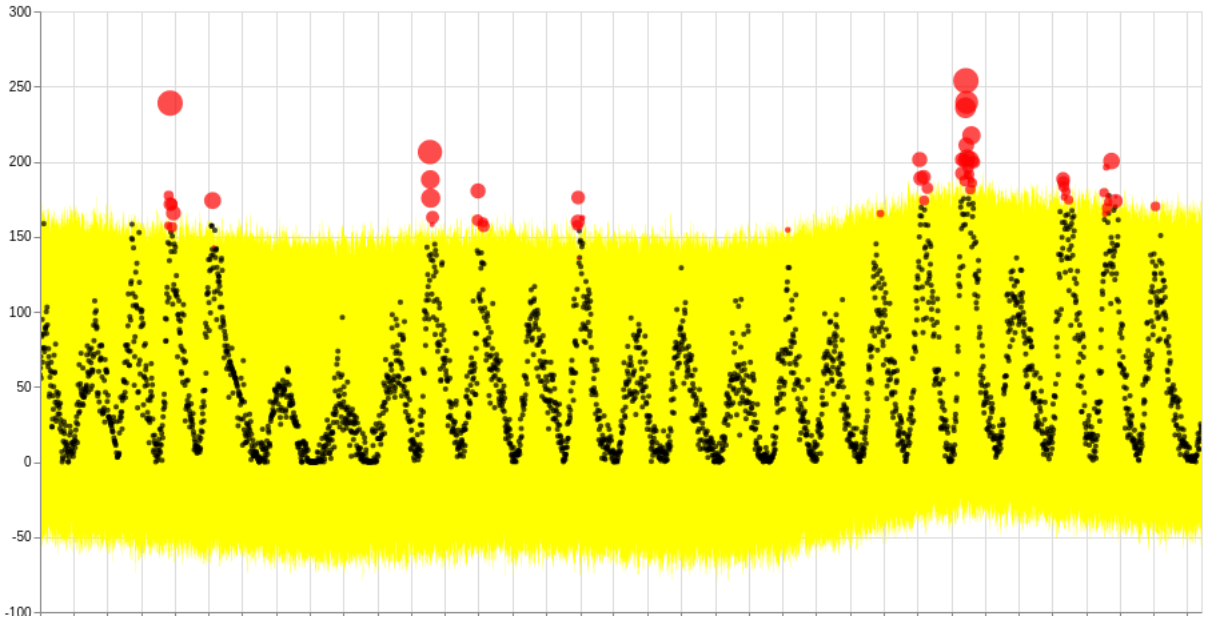

Figure 4: Outliers in sample data.

3.3. Pattern Recognition

The main aim of pattern recognition in FENIX is to provide the users and the DSS with sets of identifiable behaviour that can be linked to specific processes, inputs and/or outputs. For example, if certain patterns of sensor data emerge whenever a certain amount of impurities is present in a material, then the user and/or the DSS may discern that, whenever these patterns occur, then the amount of impurities is above this amount with high probability.

The pattern recognition module is heavily dependent upon both the amount and the characteristics of data that will be produced in the course of the pilot execution. Therefore, it has to be versatile in terms of configuration as to be able to handle various patterns that may emerge from data processing. In a similar fashion with Outlier Detection, the patterns are not known beforehand. Therefore, autoregression models have been used that may identify and extract patterns without the need of labelling them in the train data. Since this may change during pilot execution, the module also includes methods for implementing Conditional Random Field (CRF). This may produce better results, if, together with using the expertise of the pilots, the data produced are correctly labelled.

For providing a proof of concept (PoC), data outside FENIX were used. ECG (Electrocardiography) data, by their nature, have a behaviour that exhibits almost identical repetitive patterns. The results of applying Pattern Recognition to such a dataset is depicted in Figure 5.

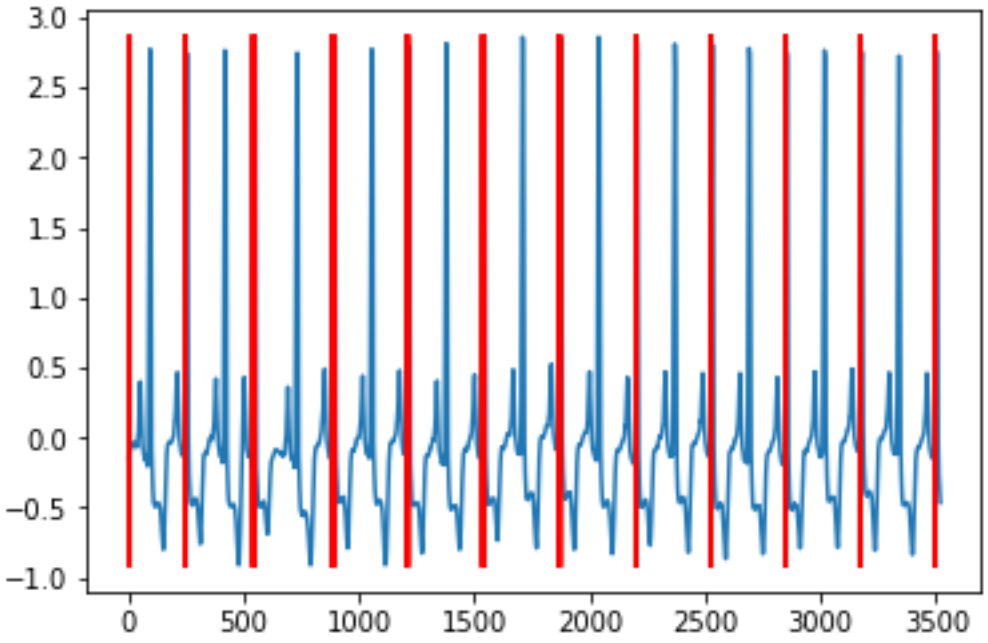

Figure 5: Pattern recognition in ECG data

4. INTEGRATION WITH THE DSS

As mentioned, the Analytics Services can be used both as standalone applications and with conjunction with the integrated DSS platform to provide results in the unified graphical environment of the DSS. As standalone, they can be run via a python shell. The output is in JSON format and plots can be exported in PNG format. Integration with the DSS platform is done via REST endpoints; the JSON output can then be rendered in the platform using an appropriate library.

Figure 6 depicts the basic flows of the Analytics Services and the integration with the DSS. As mentioned, interaction can occur either directly through an interface exposed by Jupyter, or through a REST API for integration to the DSS Platform.

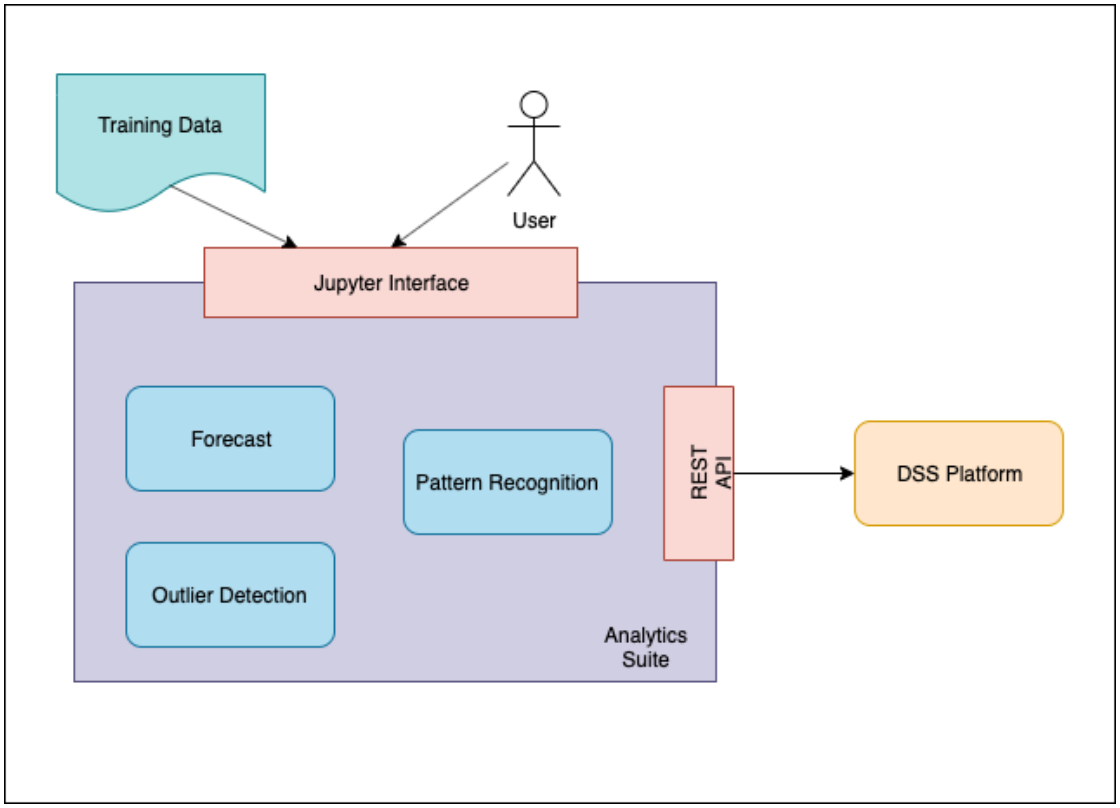

Figure 6: DSS Analytics: Architecture and integration with the DSS

5. CONCLUSION

The present document gives an overview of the developed suite of methodologies that is to be used in the context of FENIX to extract useful insight from the data; insight that can drive the decision making process both by its own or in conjunction with the various modules of the integrated DSS Engine (e.g. the Rule Engine). The modules can be connected to both streaming and batch data and are therefore ready for deployment as soon as the pilot activities of WP6 starts to provide results.
The modules developed for the DSS Analytics Services Suite incorporate many states of the art techniques in machine learning and artificial intelligence (AI). The drawback for many of these methods (e.g. neural networks), is that their exact design depends upon the dataset. While the main broad characteristics of the dataset can be discerned from the outputs of WP1 and WP3, the exact data will only be available from WP6 activities. Many algorithms work well with any dataset in an unsupervised mode. However, the best results can be in general obtained with supervised algorithms

that will be trained in the datasets produced with annotations and feedback provided by the experts. To this end, the modules support both modes and are configurable in terms of the algorithm to be produced. This knowledge, combined with the knowledge produced during WP6 activities will lead to the improvement of the quality of predictions and conclusions, that the algorithms of the DSS Analytics Module will provide.

REFERENCES

[1] Shumway, R. H., & Stoffer, D. S. (2010). Time Series Analysis and Its Applications. Springer.