

# Requirements Document for KEA

J. Cascitti      Y. Reinhart      A. Wieprecht      A. Zerbe

November 12, 2015



# Glossary

**Apache Tomcat** Apache Tomcat, often referred to as Tomcat, is an open-source web server developed by the Apache Software Foundation (ASF). Tomcat implements several Java EE specifications including Java Servlet, JavaServer Pages (JSP), Java EL, and WebSocket, and provides a "pure Java" HTTP web server environment for Java code to run in.. 29

**garbage collection** Garbage collection is a routine to search unused files and data sets. Found files will be deleted. 23

**Hyper Text Markup Language 5** HTML 5 is a markup language which is used for structuring and presenting content for the World Wide Web. 26

**Hypertext Transfer Protocol Secure** Hypertext Transfer Protocol Secure (HTTPS) is a communications protocol for secure communication over a computer network, with especially wide deployment on the Internet. Technically, it is not a protocol in and of itself; rather, it is the result of simply layering the Hypertext Transfer Protocol (HTTP) on top of the SSL/TLS protocol, thus adding the security capabilities of SSL/TLS to standard HTTP communications. (source: en.wikipedia.org: 11.11.2013). 25, 31

**J48** J48 is an open source Java implementation of the C4.5 decision tree algorithm. It creates a decision tree.. 34

**Java** Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation. As of 2015, Java is one of the most popular programming languages in use, particularly for client-server web applications. The latest version is Java 8. ([https://en.wikipedia.org/wiki/Java\\_%28programming\\_language%29](https://en.wikipedia.org/wiki/Java_%28programming_language%29)). 26, 29

**Java servlets** A Java servlet is a Java program that extends the capabilities of a server. Although servlets can respond to any types of requests, they

most commonly implement applications hosted on Web servers. Such Web servlets are the Java counterpart to other dynamic Web content technologies such as PHP. 29, 30

**JavaScript** JavaScript is an interpreted computer programming language. As part of web browsers, implementations allow client-side scripts to interact with the user, control the browser, communicate asynchronously, and alter the document content that is displayed. It has also become common in server-side programming, game development, and the creation of desktop applications. (source: en.wikipedia.org : 11.11.2013). 26

**Plugins** A plugin is an extension to any sort of software that adds additional features or alters existing ones. 8, 33, 36

**Random forest** Random forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forests correct for decision trees' habit of overfitting to their training set.. 34

**Representational state Transfer (REST)** The Representational State Transfer (REST) is a software architectural style. A system that is conform to the constraints of REST can be called RESTful. That means, that the system will communicate over the Hypertext Transfer Protocol with the same HTTP verbs that browsers use for their communication with web servers. Resources are addressed through distinct URI and the system is based on a client-server model.. 29

**Resource Description Framework** RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a *triple*). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications. This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations. (<http://www.w3.org/RDF/>). 29

**SPARQL** SPARQL (pronounced "sparkle") is an RDF query language, that is, a semantic query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.. 22

# Contents

Glossary	3
----------	---

<b>I</b>	<b>User Requirements</b>	<b>9</b>
0.0.0.0.1	UR001 . . . . .	11
0.0.0.0.2	UR002 . . . . .	11
0.0.0.0.3	UR003 . . . . .	11
0.0.0.0.4	UR004 . . . . .	11
0.0.0.0.5	UR005 . . . . .	11
0.0.0.0.6	UR006 . . . . .	11
0.0.0.0.7	UR007 . . . . .	11
0.0.0.0.8	UR008 . . . . .	11
0.0.0.0.9	UR009 . . . . .	12
0.0.0.0.10	UR010 . . . . .	12
0.0.0.0.11	UR011 . . . . .	12
0.0.0.0.12	UR012 . . . . .	12
0.0.0.0.13	UR013 . . . . .	12
0.0.0.0.14	UR014 . . . . .	12
0.0.0.0.15	UR015 . . . . .	12
0.0.0.0.16	UR016 . . . . .	12
0.0.0.0.17	UR017 . . . . .	13
0.0.0.0.18	UR018 . . . . .	13
0.0.0.0.19	UR019 . . . . .	13
0.0.0.0.20	UR020 . . . . .	13
0.0.0.0.21	UR021 . . . . .	13
0.0.0.0.22	UR022 . . . . .	13
0.0.0.0.23	UR023 . . . . .	13
0.0.0.0.24	UR024 . . . . .	13
0.0.0.0.25	UR025 . . . . .	14
0.0.0.0.26	UR026 . . . . .	14
0.0.0.0.27	UR027 . . . . .	14
0.0.0.0.28	UR028 . . . . .	14
0.0.0.0.29	UR029 . . . . .	14
0.0.0.0.30	UR030 . . . . .	14

0.0.0.0.31	UR031	14
0.0.0.0.32	UR032	15
0.0.0.0.33	UR033	15
0.0.0.0.34	UR034	15
0.0.0.0.35	UR035	15
0.0.0.0.36	UR036	15
0.0.0.0.37	UR037	15
0.0.0.0.38	UR038	15
0.0.0.0.39	UR039	16
0.0.0.0.40	UR040	16
0.0.0.0.41	UR041	16
0.0.0.0.42	UR042	16
0.0.0.0.43	UR043	16
0.0.0.0.44	UR044	16
0.0.0.0.45	UR045	16
0.0.0.0.46	UR046	16
0.0.0.0.47	UR047	17
0.0.0.0.48	UR048	17
0.0.0.0.49	UR049	17
0.0.0.0.50	UR050	17
0.0.0.0.51	UR051	17
0.0.0.0.52	UR052	17
0.0.0.0.53	UR053	17
0.0.0.0.54	UR054	17
0.0.0.0.55	UR055	18
0.0.0.0.56	UR056	18
0.0.0.0.57	UR057	18
0.0.0.0.58	UR058	18
0.0.0.0.59	UR059	18
0.0.0.0.60	UR060	18
0.0.0.0.61	UR061	18
0.0.0.0.62	UR062	18

## II System Requirements 19

### 1 Non-Functional Requirements 21

#### 1.1 Product Requirements 21

##### 1.1.1 Usability Requirements 21

###### 1.1.1.0.1 NFR001 21

###### 1.1.1.0.2 NFR002 21

###### 1.1.1.0.3 NFR003 21

###### 1.1.1.0.4 NFR004 21

###### 1.1.1.0.5 NFR005 22

###### 1.1.1.0.6 NFR006 22

###### 1.1.1.0.7 NFR007 22

1.1.1.0.8	NFR008 . . . . .	22
1.1.1.0.9	NFR009 . . . . .	22
1.1.1.0.10	NFR010 . . . . .	22
1.1.1.0.11	NFR011 . . . . .	22
1.1.1.0.12	NFR012 . . . . .	23
1.1.1.0.13	NFR013 . . . . .	23
1.1.2	Efficiency Requirements . . . . .	23
1.1.2.1	Performance Requirements . . . . .	23
1.1.2.1.1	NFR014 . . . . .	23
1.1.2.1.2	NFR015 . . . . .	23
1.1.2.1.3	NFR016 . . . . .	23
1.1.2.2	Space Requirements . . . . .	23
1.1.2.2.1	NFR017 . . . . .	23
1.1.3	Dependability Requirements . . . . .	24
1.1.3.0.1	NFR018 . . . . .	24
1.1.3.0.2	NFR019 . . . . .	24
1.1.3.0.3	NFR020 . . . . .	24
1.1.3.0.4	NFR021 . . . . .	24
1.1.3.0.5	NFR022 . . . . .	24
1.1.3.0.6	NFR023 . . . . .	24
1.1.3.0.7	NFR024 . . . . .	24
1.1.4	Security Requirements . . . . .	25
1.1.4.0.1	NFR025 . . . . .	25
1.1.4.0.2	NFR026 . . . . .	25
1.1.4.0.3	NFR027 . . . . .	25
1.2	Organisational Requirements . . . . .	25
1.2.1	Environmental Requirements . . . . .	25
1.2.1.0.1	NFR028 . . . . .	25
1.2.1.0.2	NFR029 . . . . .	25
1.2.1.0.3	NFR030 . . . . .	25
1.2.1.0.4	NFR031 . . . . .	25
1.2.1.0.5	NFR032 . . . . .	26
1.2.1.0.6	NFR033 . . . . .	26
1.2.2	Operational Requirements . . . . .	26
1.2.2.0.1	NFR034 . . . . .	26
1.2.3	Development Requirements . . . . .	26
1.2.3.0.1	NFR035 . . . . .	26
1.2.3.0.2	NFR036 . . . . .	26
1.2.3.0.3	NFR037 . . . . .	26
1.3	External Requirements . . . . .	26
1.3.1	Regulatory Requirements . . . . .	26
1.3.1.0.1	NFR038 . . . . .	26
1.3.1.0.2	NFR039 . . . . .	27
1.3.2	Ethical Requirements . . . . .	27
1.3.3	Legislative Requirements . . . . .	27
1.3.3.1	Accounting Requirements . . . . .	27

1.3.3.2	Safety / Security Requirements . . . . .	27
1.3.3.2.1	NFR040 . . . . .	27
1.3.3.2.2	NFR041 . . . . .	27
<b>2</b>	<b>Functional Requirements</b>	<b>29</b>
2.0.0.0.1	FR001 . . . . .	29
2.0.0.0.2	FR002 . . . . .	29
2.0.0.0.3	FR003 . . . . .	29
2.0.0.0.4	FR004 . . . . .	29
2.0.0.0.5	FR005 . . . . .	29
2.0.0.0.6	FR006 . . . . .	30
2.0.0.0.7	FR007 . . . . .	30
2.0.0.0.8	FR008 . . . . .	30
2.0.0.0.9	FR009 . . . . .	30
2.0.0.0.10	FR010 . . . . .	30
2.0.0.0.11	FR011 . . . . .	30
2.0.0.0.12	FR012 . . . . .	31
2.0.0.0.13	FR013 . . . . .	31
2.0.0.0.14	FR014 . . . . .	31
2.0.0.0.15	FR015 . . . . .	31
2.0.0.0.16	FR016 . . . . .	31
2.0.0.0.17	FR017 . . . . .	31
2.0.0.0.18	FR018 . . . . .	32
2.0.0.0.19	FR019 . . . . .	32
2.0.0.0.20	FR020 . . . . .	32
2.0.0.0.21	FR021 . . . . .	32
2.0.0.0.22	FR022 . . . . .	32
2.0.0.0.23	FR023 . . . . .	32
2.0.0.0.24	FR024 . . . . .	32
2.0.0.0.25	FR025 . . . . .	33
2.0.0.0.26	FR026 . . . . .	33
2.0.0.0.27	FR027 . . . . .	33
2.0.0.0.28	FR028 . . . . .	33
2.0.0.0.29	FR029 . . . . .	33
2.0.0.0.30	FR030 . . . . .	33
2.0.0.0.31	FR031 . . . . .	33
2.0.0.0.32	FR032 . . . . .	34
2.0.0.0.33	FR033 . . . . .	34
2.0.0.0.34	FR034 . . . . .	34
<b>3</b>	<b>Scenarios</b>	<b>35</b>
3.1	Logging in to the system . . . . .	35
3.2	Upload data . . . . .	35
3.3	Upload algorithm Plugins . . . . .	36
3.4	Training a model . . . . .	37
3.5	Testing data . . . . .	37



# Part I

## User Requirements



**0.0.0.0.1 UR001**

**Statement** *The software shall be published under an open source license.*

**Priority** *A*

**0.0.0.0.2 UR002**

**Statement** *There shall be no monetarization of the software, such as advertisements.*

**Priority** *A*

man benennt nicht  
unbedingt das, was nicht  
sein soll

**0.0.0.0.3 UR003**

**Statement** *The language of the website shall be English.*

**Priority** *A*

**0.0.0.0.4 UR004**

**Statement** *The software shall be developed in the Java programming language.*

**Priority** *A*

**0.0.0.0.5 UR005**

**Statement** *The target audience is IT personnel, mainly data analysts.*

**Priority** *A*

**0.0.0.0.6 UR006**

**Statement** *The average user shall be able to learn the operation of the system in one day.*

**Priority** *A*

**0.0.0.0.7 UR007**

**Statement** *The administrators shall be able to learn the operation of the system in ten days.*

**Priority** *A*

war nicht als UR?!  
Zeitspanne eher in  
NFR! -->Bsp  
wurden nur  
erwähnt.  
Eher: "System soll  
intuitiv sein" (UR) --  
> "erlernbar in 10  
std"

**0.0.0.0.8 UR008**

**Statement** *The system shall be expendable after deployment.*

**Priority** *A*

**0.0.0.0.9 UR009**

**Statement** *The software shall be accessible with all common webbrowsers.*

**Priority** *A*

**0.0.0.0.10 UR010**

**Statement** *The website should offer a version for mobile devices.*

**Priority** *B*

**0.0.0.0.11 UR011**

**Statement** *The system should be accessible through a command line interface via SPARQL.*

**Priority** *A*

**0.0.0.0.12 UR012**

**Statement** *The system shall accept standardized data types accepted by Weka for data upload by the user.*

**Priority** *A*

**0.0.0.0.13 UR013**

**Statement** *The system shall run on desktop environments. Already existing server infrastructure provides 32GB of RAM and Octa-Core CPUs.*

**Priority** *A*

zu speziell für UR?!

**0.0.0.0.14 UR014**

**Statement** *The only required form of encryption of communication between user and software is https.*

**Priority** *A*

**0.0.0.0.15 UR015**

keine benennung  
desse, was nicht ist

**Statement** *Caching of results is not required. All required calculations shall be performed for each user request on the server.*

**Priority** *A*

**0.0.0.0.16 UR016**

**Statement** *The system shall provide interfaces for external plugins.*

**Priority** *C*

Priorität C?

**0.0.0.0.17 UR017****Statement** *User passwords shall be stored in the form of salted hashes.***Priority** *A*

zu speziell? wurde  
das so genannt?  
oder ist das schon  
SR und  
Umsetzung?

**0.0.0.0.18 UR018****Statement** *The system shall not have its own backup system.***Priority** *A*

Nennung desse,  
was nicht enthalten

**0.0.0.0.19 UR019****Statement** *The server infrastructure shall run on a GNU/Linux operating system.***Priority** *A***0.0.0.0.20 UR020****Statement** *The system infrastructure does not need to be redundant.***Priority** *A*

nicht enthalten

**0.0.0.0.21 UR021****Statement** *Data sets, models and result sets shall be organized in packages which are each assigned a unique ID.***Priority** *A***0.0.0.0.22 UR022****Statement** *Package contents shall be accessible through their ID.***Priority** *A***0.0.0.0.23 UR023****Statement** *Different installations of the system shall be able to exchange models.***Priority** *A*

?

**0.0.0.0.24 UR024****Statement** *If there is a problem with the server the system shall reboot. There is no server to sidestep.***Priority** *A*

nicht enthalten...  
und zu speziell?

**0.0.0.0.25 UR025**

**Statement** *If a data upload is interrupted, it shall be restarted from the beginning.*

**Priority** *A*

**0.0.0.0.26 UR026**

**Statement** *The software shall offer a user account system with different user groups which possess different rights, such as administrators, normal users and an anonymous user account for unregistered users.*

**Priority** *A*

**0.0.0.0.27 UR027**

welche algorithmen?

**Statement** *Users shall be able to flexibly choose algorithms to be applied to a dataset of their choice.*

**Priority** *A*

**0.0.0.0.28 UR028**

**Statement** *Settings chosen by a user shall be saved, independently for each user.*

**Priority** *A*

**0.0.0.0.29 UR029**

**Statement** *A history of user activity is not necessary.*

nicht können..

**Priority** *A*

**0.0.0.0.30 UR030**

**Statement** *Further processing of results, i.e. forms of display such as trees, graphs, shall be selectable flexibly by the user.*

**Priority** *A*

**0.0.0.0.31 UR031**

**Statement** *Algorithms uploaded by the user must be approved by an administrator before being selectable by a user.*

**Priority** *A*

**0.0.0.0.32 UR032**

**Statement** *The administrators shall have a control menu to modify user rights and system preferences.*

**Priority** *A*

**0.0.0.0.33 UR033**

**Statement** *The software shall, with sensible restraints, be accessible without logging in as a user.*

**Priority** *A*

**0.0.0.0.34 UR034**

**Statement** *The software shall save queries and results per user, so they can be accessed in the future.*

**Priority** *A*

**0.0.0.0.35 UR035**

**Statement** *Visibility of results and queries to other users shall be configurable by the applicant.*

...außer bei  
anonymen Nutzern

**Priority** *A*

**0.0.0.0.36 UR036**

**Statement** *There shall be no messaging system for users, such as for private messages between users.*

nicht können...

**Priority** *A*

**0.0.0.0.37 UR037**

**Statement** *Data which can be uploaded by users, such as datasets and learned models, shall also be downloadable by the user.*

**Priority** *A*

**0.0.0.0.38 UR038**

**Statement** *While the run time for registered users is unlimited, the system shall offer users the option to stop an algorithm.*

**Priority** *A*

es gab doch aber  
keine memory  
constraints oder?

**0.0.0.0.39 UR039**

**Statement** *Users shall be notified when the system's storage or memory are full.*

**Priority** *A*

**0.0.0.0.40 UR040**

**Statement** *Users shall be able to delete data uploaded by them.*

**Priority** *A*

**0.0.0.0.41 UR041**

**Statement** *Anonymous user packages shall be deleted after a certain time.*

**Priority** *A*

...nicht können.  
Besser  
umformulieren in  
das, was der nicht  
registrierte user  
nicht kann

**0.0.0.0.42 UR042**

**Statement** *The number of data sets a signed-in user can upload is not restricted.*

**Priority** *A*

welche terms und  
conditions?

**0.0.0.0.43 UR043**

**Statement** *The terms and conditions shall be changeable by the administrator.*

**Priority** *A*

**0.0.0.0.44 UR044**

**Statement** *A user shall be able to decide whether to let the system divide a data set into training and test data sets or upload them separately.*

**Priority** *A*

**0.0.0.0.45 UR045**

**Statement** *A user shall be able to upload a learned model.*

**Priority** *A*

**0.0.0.0.46 UR046**

**Statement** *It shall be possible to add new algorithms to be offered to the user to the system.*

**Priority** *A*



**0.0.0.0.47 UR047**

**Statement** *The implemented algorithms offered to the user shall be extensible and modifiable.*

eher priorität B?

**Priority** *A*

**0.0.0.0.48 UR048**

**Statement** *An editor for algorithms, accesible to the user, is not required. Instead users shall be able to upload .jar-files to add own algorithms to be used on datasets to the system.*

prio B?!

**Priority** *A*

**0.0.0.0.49 UR049**

**Statement** *The used database shall be an RDF database.*

**Priority** *A*

**0.0.0.0.50 UR050**

**Statement** *The system shall focus on classification algorithms.*

**Priority** *A*

**0.0.0.0.51 UR051**

**Statement** *The system shall support the algorithms SMO, Random Forest and J48.*

**Priority** *A*

**0.0.0.0.52 UR052**

**Statement** *The system shall offer clustering algorithms.*

**Priority** *C*

warum nicht Prio B? Und wenn C, dann nicht "shall" benutzen..

**0.0.0.0.53 UR053**

**Statement** *The algorithms shall be executed wholly on the server.*

**Priority** *A*

**0.0.0.0.54 UR054**

**Statement** *Each algorithm's required and optional parameters shall be presented to the user.*

**Priority** *A*

**0.0.0.0.55 UR055**

**Statement** *The presentation of the results shall be dependent on the used algorithm.*

**Priority** *B*

**0.0.0.0.56 UR056**

**Statement** *The system shall provide restrictions to stop the computation of too many algorithms at the same time.*

**Priority** *A*

hier fehlt die  
bevorzugung der  
reg. Users?

**0.0.0.0.57 UR057**

**Statement** *Data sets shall be partitionable after upload.*

**Priority** *A*

partitionierbar? für  
was? in test und  
data sets?!

**0.0.0.0.58 UR058**

**Statement** *For each algorithm there shall be an own input and an output mask.*

**Priority** *A*

maske?

**0.0.0.0.59 UR059**

**Statement** *A learned model cannot be changed.*

**Priority** *A*

**0.0.0.0.60 UR060**

**Statement** *The result of the algorithm shall be downloadable in several formats.*

**Priority** *A*

**0.0.0.0.61 UR061**

**Statement** *The raw data sets shall be downloadable.*

**Priority** *A*

**0.0.0.0.62 UR062**

**Statement** *Unregistered users shall have limited algorithm run time.*

**Priority** *A*

# Part II

## System Requirements



# Chapter 1

## Non-Functional Requirements

### 1.1 Product Requirements

#### 1.1.1 Usability Requirements

##### 1.1.1.0.1 NFR001

**Statement** *After reading a short text introduction about the main features (upload and processing of .arff data, creating models, creating custom algorithms, receiving and sharing results) of the system, common users should be able to use the main features of this software. The average user shall be able to learn the operation of the system in one day.*

**Priority** *A*

Bezug zu UR ?

##### 1.1.1.0.2 NFR002

**Statement** *It shall be provided a documentation for advanced users which explains all features.*

**Priority** *A*

##### 1.1.1.0.3 NFR003

**Statement** *It shall be provided a documentation for administrators which explains all features and common operating cycles. The administrators shall be able to learn the operation of the system in ten days.*

**Priority** *A*

##### 1.1.1.0.4 NFR004

**Statement** *The system can be accessed and used if there is an internet connection. There are no restrictions regarding firewalls and routers.*

**Priority** *A*

letzter satz unnötig?

**1.1.1.0.5 NFR005**

**Statement** *The application should be usable for following browsers, since the listed versions: Firefox 35, Chrome 40, Opera 27, Internet Explorer 11 and Microsoft Edge 20.*

**Priority** *A*

**1.1.1.0.6 NFR006**

**Statement** *The website should provide a dedicated view for mobile browsers, like Safari Mobile, Chrome Mobile and Dolphin.*

**Priority** *B*

**1.1.1.0.7 NFR007**

**Statement** *The system language is English.*

**Priority** *A*

**1.1.1.0.8 NFR008**

**Statement** *The targeted audience is IT personnel, mainly data analysts. Detailed options are preferred instead of oversimplified user interfaces.*

**Priority** *A*

**1.1.1.0.9 NFR009**

**Statement** *The system should be accessible through a command line interface via SPARQL.*

**Priority** *A*

**1.1.1.0.10 NFR010**

**Statement** *If a data upload is interrupted, it shall be restarted from the beginning.*

**Priority** *A*

NFRs sind gleich den USR?!

**1.1.1.0.11 NFR011**

**Statement** *The software has no integrated system for users to send messages to eachother.*

**Priority** *A*

sind das überhaupt  
noch Usability  
Requirements?!

**1.1.1.0.12 NFR012**

**Statement** *The implemented algorithm shall include mainly classification algorithms.*

**Priority** *A*

**1.1.1.0.13 NFR013**

**Statement** *The implemented algorithms shall also include clustering algorithms.*

**Priority** *B*

**1.1.2 Efficiency Requirements****1.1.2.1 Performance Requirements****1.1.2.1.1 NFR014**

**Statement** *The minimal system requirements demand a desktop environment or server with following specifications: i5 CPU with 8GB Ram, 500 GB storage, and a GNU/Linux operating system.*

**Priority** *A*

**1.1.2.1.2 NFR015**

**Statement** *The recommended system requirements for the server are as follows: Intel Xeon CPU with 32GB Ram, 2 TB storage, and a GNU/Linux operating system.*

**Priority** *A*

**1.1.2.1.3 NFR016**

**Statement** *At most, two algorithms shall be executed at a time. All further requests by users must be enqueued.*

**Priority** *A*

**1.1.2.2 Space Requirements****1.1.2.2.1 NFR017**

**Statement** *Once a day garbage collection is triggered to search expired guest accounts, session recordings, and shared files.*

Expired files are defined as follows:

**guest data** After 30 days.

**registered user data** Never.

**Priority** *A*

### 1.1.3 Dependability Requirements

#### 1.1.3.0.1 NFR018

**Statement** *If the system crashes, the system shall try to reboot automatically and inform the administrator.*

**Priority** *A*

#### 1.1.3.0.2 NFR019

**Statement** *If the system crashes, there is no server to sidestep.*

**Priority** *A*

#### 1.1.3.0.3 NFR020

**Statement** *The system has no integrated backup system.*

**Priority** *A*

#### 1.1.3.0.4 NFR021

**Statement** *The system shall be available 24/7.*

**Priority** *A*

#### 1.1.3.0.5 NFR022

**Statement** *The number of system crashes in a month should be less than one.*

**Priority** *B*

#### 1.1.3.0.6 NFR023

**Statement** *Full functionality of the system shall be restored 30 minutes after a system crash.*

**Priority** *B*

#### 1.1.3.0.7 NFR024

**Statement** *The downtime during regular work hours shall not exceed 30 minutes per day.*

**Priority** *B*



### 1.1.4 Security Requirements

#### 1.1.4.0.1 NFR025

**Statement** *Data transfer to and from the server must be performed via an Hypertext Transfer Protocol Secure connection.*

**Priority** *A*

#### 1.1.4.0.2 NFR026

**Statement** *Passwords shall be stored in the form of salted hashes.*

**Priority** *A*

#### 1.1.4.0.3 NFR027

**Statement** *All data is stored unencrypted on the server.*

**Priority** *A*

## 1.2 Organisational Requirements

### 1.2.1 Environmental Requirements

#### 1.2.1.0.1 NFR028

**Statement** *Users should identify themselves by a name, e-mail address and password.*

**Priority** *A*

#### 1.2.1.0.2 NFR029

**Statement** *Each registered user belongs to one of the following user groups: normal user, administrator.*

**Priority** *A*

#### 1.2.1.0.3 NFR030

**Statement** *Apart from the basic user groups, user can organize themselves in groups, which can be given read and write rights by data owners.*

**Priority** *A*

#### 1.2.1.0.4 NFR031

**Statement** *Unregistered users share an anonymous user account.*

**Priority** *A*

hier sollte es doch  
eher um  
Unternehmensanfor-  
derungen gehen  
bez. operationale  
Prozessanforderun-  
gen,  
Entwicklungsproze-  
sse etc.?

**1.2.1.0.5 NFR032**

**Statement** *Preferences set by users are saved independently for each user. The anonymous is locked into default settingsi set by an administrator.*

**Priority** *A*

**1.2.1.0.6 NFR033**

**Statement** *A history of user activity will not be created.*

...nicht können

**Priority** *A*

**1.2.2 Operational Requirements****1.2.2.0.1 NFR034**

**Statement** *There will be no monetarization of the software, such as advertisements.*

**Priority** *A*

**1.2.3 Development Requirements****1.2.3.0.1 NFR035**

**Statement** *The software shall be developed in the Java programming language.*

**Priority** *A*

**1.2.3.0.2 NFR036**

**Statement** *The web interface shall be valid Hyper Text Markup Language 5, CSS and JavaScript.*

**Priority** *A*

**1.2.3.0.3 NFR037**

**Statement** *The system shall be expendable after deployment.*

**Priority** *A*

**1.3 External Requirements****1.3.1 Regulatory Requirements****1.3.1.0.1 NFR038**

**Statement** *The system is published under the GPL Version 3.*

GPL`?

**Priority** *A*

**1.3.1.0.2 NFR039**

**Statement** *Administrators have the capability and responsibility to adjust or update the terms and condition.*

**Priority** *A*

**1.3.2 Ethical Requirements**

**1.3.3 Legislative Requirements**

**1.3.3.1 Accounting Requirements**

**1.3.3.2 Safety / Security Requirements**

**1.3.3.2.1 NFR040**

**Statement** *The system respects German law.*

**Priority** *A*

**1.3.3.2.2 NFR041**

**Statement** *Storing data of members shall be conform the German data legislations.*

**Priority** *A*



## Chapter 2

# Functional Requirements

### 2.0.0.0.1 FR001

**Statement** *The system utilises an Resource Description Framework database to store and organize all data. (see user requirement UR049)*

**Priority** *A*

### 2.0.0.0.2 FR002

**Statement** *Learned models are saved as serialized Java objects.*

referenzen?

**Priority** *A*

### 2.0.0.0.3 FR003

**Statement** *The web interface is based on Representational state Transfer (REST), Apache Tomcat and Java servlets.*

**Priority** *A*

### 2.0.0.0.4 FR004

**Statement** *User can registers on the server.*

?

**Prioriy** *A*

### 2.0.0.0.5 FR005

**Statement** *User can use the system without registering as a user via a shared anonymous user account that utilises default settings set by an administrator. (see user requirement UR033)*

**Prioriy** *A*

welche genauen settings?

**2.0.0.0.6 FR006**

**Statement** *The system makes use of Java servlets to upload data sets as .arff or csv files. Other data types shall not be allowed. If the user tries to upload a wrong data type, a message shall be displayed. (see user requirement UR012)*

**Priority** *A*

algorithmen  
abhängig von den  
daten?

**2.0.0.0.7 FR007**

**Statement** *After selecting a dataset, users will see a menu from which they will select the algorithm they want to use. (see user requirement UR027)*

**Priority** *A*

**2.0.0.0.8 FR008**

**Statement** *When selecting a data set and an algorithm to be applied to it, user can select which fraction of the data set shall be used as training data and test data respectively or select a different data set for each role. (see user requirement UR044)*

**Priority** *A*

**2.0.0.0.9 FR009**

**Statement** *When selecting a data set and an algorithm to be applied to it, the user is presented input fields for the specific parameters of the selected algorithm. These are tagged as either optional and required.*

**Priority** *A*

**2.0.0.0.10 FR010**

**Statement** *The system allows to set restrictions to visibility of results and queries to other users. (see user requirement UR035)*

**Priority** *A*

**2.0.0.0.11 FR011**

**Statement** *The user menu allows the user to set read and write rights for data sets, models, results or whole packages owned by them to user groups.*

**Priority** *A*

**2.0.0.0.12 FR012**

**Statement** *User can choose to download raw datasets, learned models and result sets in different formats. The format selection is filtered to only include appropriate formats for the given data. (see user requirements UR061, UR062)*

**Priority** *A*

**2.0.0.0.13 FR013**

**Statement** *User can choose to upload raw datasets and learned models. (see user requirement UR045)*

**Priority** *A*

**2.0.0.0.14 FR014**

**Statement** *Data sets, models and saved results can only be deleted by their respective owner (i.e. the user who originally uploaded them/requested their creation by the system), administrators and those users who have been given write-rights by the owner. (see user requirement UR040)*

**Priority** *A*

**2.0.0.0.15 FR015**

**Statement** *The system allows to upload completely new algorithms as .jar file into the custom algorithm section. (see user requirement UR046)*

Priorität B?!

**Prioriy** *A*

**2.0.0.0.16 FR016**

**Statement** *All kinds of communication through WAN or LAN shall be encrypted by the Hypertext Transfer Protocol Secure standard. (see user requirement UR014)*

**Priority** *B*

**2.0.0.0.17 FR017**

**Statement** *The system must give the opportunity of user-registration via e-mail-address, username, and password. (see user requirement UR026)*

**Priority** *A*

**2.0.0.0.18 FR018**

wie schaut das  
genau aus?

**Statement** *The system supports registered and unregistered users. (see user requirement UR026)*

**Priority** *A*

**2.0.0.0.19 FR019**

**Statement** *Registered shall be able to store their settings made at previous logins. Unregistered users always start with default settings and changes they made are only valid for their current session. (see user requirement UR034)*

**Priority** *A*

**2.0.0.0.20 FR020**

**Statement** *Anonymous user packages will be deleted after 30 days. (see user requirement UR041)*

**Priority** *A*

**2.0.0.0.21 FR021**

**Statement** *All required calculations shall be performed for each user request. No caching of results. (see user requirement UR015)*

**Priority** *A*

**2.0.0.0.22 FR022**

**Statement** *All required calculations shall be performed on the server. (see user requirement UR053)*

**Priority** *A*

**2.0.0.0.23 FR023**

**Statement** *The calculation run time for registered users is unlimited. (see user requirement UR038)*

**Priority** *A*

**2.0.0.0.24 FR024**

**Statement** *The number of data sets a registered user can upload is not restricted. (see user requirement UR042)*

**Priority** *A*

only the number or  
is it depending on  
the size



**2.0.0.0.25 FR025**

**Statement** *Users can abort running algorithms at any given time with a stop button. (see user requirement UR038)*

**Priority** *A*

**2.0.0.0.26 FR026**

**Statement** *The algorithm run time for the anonymous user account is limited. The time limit can be set by administrators. (see user requirement UR063)*

**Priority** *A*

**2.0.0.0.27 FR027**

**Statement** *The system shall provide interfaces for external Plugins. (see user requirement UR016)*

**Priority** *A*

**2.0.0.0.28 FR028**

**Statement** *The system shall organize data sets, models and result sets in packages which have a unique ID. (see user requirement UR021)*

**Priority** *A*

**2.0.0.0.29 FR029**

**Statement** *Packages and their contents shall be accessible through the package's ID. (see user requirement UR022)*

**Priority** *A*

**2.0.0.0.30 FR030**

**Statement** *Data packages possess an address that identifies their location, including which server they are located on. Through this address, users can access data on different installations of the system. (see user requirement UR023)*

**Priority** *A*

**2.0.0.0.31 FR031**

**Statement** *After creation of results the user will be offered a selection of appropriate forms of presentation, such as graphs, trees, etc. (see user requirement UR030)*

**Priority** *A*

**2.0.0.0.32 FR032**

**Statement** *The administrator interface allows to approve custom algorithms uploaded by users, making them selectable by users. (see user requirement UR031)*

**Priority** *A*

**2.0.0.0.33 FR033**

**Statement** *The algorithms offered to the user shall include SMO, Random forest and J48. (see user requirement UR051)*

**Priority** *A*

**2.0.0.0.34 FR034**

**Statement** *The user menu allows users to edit data sets and partition them into separate sets. Learned models cannot be modified.*

**Priority** *A*

## Chapter 3

# Scenarios

### 3.1 Logging in to the system

**INITIAL ASSUMPTION:** *The user has an internet browser running and opened the system website.*

**NORMAL:** *The user clicks on the login-button located at a prominent position on the website. A dialogue window opens up and the user has to fill his e-mail-address and password into the form. After submitting the systems window interface opens up in the browserwindow. Initially there are no plugins open. If the user had active plugins during his last logout, these plugins will be opened again.*

**WHAT CAN GO WRONG:**

*The connection to the database is damaged and therefor the login information can not be confirmed. In this case there will be an error message for the user.*

*The user types in a wrong e-mail address or a wrong password. In this case a message appears, that the e-mail address or the password is wrong. The user is not registered to the System. In this case an error message appears, that the e-mail address or the password is wrong.*

*The connection to the internet is lost.*

only a error  
massage?

**OTHER ACTIVITIES:**

**SYSTEM STATE ON COMPLETION:** *The user is logged in and able to use the system.*

### 3.2 Upload data

**INITIAL ASSUMPTION:** *The user has an internet browser running and opened the system website, where he could be logged in as registered user,*

was ist hier mit dem unregistriertem User? "logged in as registered user, but it is not necessary" ?

*but it is not necessary.*

**NORMAL:** *The user pushes a button which is named with 'upload data'. Thereupon a window appears where the user can choose the data file he wants to upload. He has to select the file by clicking on it. When he confirms his choice by clicking on an 'ok'- button the system starts to prove if the file has a permitted data type. If this is verified, the uploading process starts.*

#### WHAT CAN GO WRONG:

*The selected file is from a not accepted data type. In this case a window with an error message pops up and there is no upload possible.*

*The internet connection of the user can be lost. In this case the user has to start the whole uploading process again, when the internet connection is restored. Already uploaded parts of the data set get lost.*

*It seems that the uploading process, will not terminate. In this case the user has the possibility to interrupt the process by clicking on a 'stop'-button. If so already uploaded parts of the data set get lost.*

#### OTHER ACTIVITIES:

**SYSTEM STATE ON COMPLETION:** *The user has a data set at which he can work.*

### 3.3 Upload algorithm Plugins

**INITIAL ASSUMPTION:** *The user has an internet browser running and opened the system website, where he is logged in as registered user.*

whiteboard?

**NORMAL:** *The user starts in another window in his browser a new algorithm-plugin. Now he wants to add the whiteboard to the system. If the user is the system administrator this is very simple. He clicks on the button share in the plugin field of the algorithm. After this the plugin is added to the system and all users are able to use the algorithm on their data. If the user who wants to add the new plugin is no administrator he has to do the same steps like the administrator, but the administrator has to activate the plugin before everybody can use it.*

new algorithms sind doch Priorität B, warum dann hier als Szenario erwähnt?

#### WHAT CAN GO WRONG:

*The session can crash while adding. In this case nothing happen, what means that the plugin won't add to the session.*

#### OTHER ACTIVITIES:

**SYSTEM STATE ON COMPLETION:** *The user has added a new algorithm to the system.*

## 3.4 Training a model

**INITIAL ASSUMPTION:** *The user has an internet browser running and opened the system website, where he is not necessarily logged in as registered user. The user has already uploaded data.*

**NORMAL:** *First of all the user has to go to the submenu item 'my data' where all his uploaded data sets are listed. From this list he has to pick the one data set by clicking on it. Now this data set is marked. After the next step, clicking on the 'train model'-button a menu appears where all algorithms are listed which are compatible with the picked data. The user choose one of the offered algorithms. This opens an input mask where he can define given parameters. There is also the opportunity to choose whether all the data will be used for training or if some data shall remain for testing afterwards. Now the training can be started, by clicking on a 'start'-button. If the training process is completed the new model is displayed in the submenu 'my models'.*

als nicht  
registrierter User  
hat man nicht die  
möglichkeit?

**WHAT CAN GO WRONG:**

*There is no compatible algorithm for the picked data. In this case the user can either upload a new algorithm which fits, or choose an other dataset. The connection to the server is interrupted. In this case the user has to start the training again.*

**OTHER ACTIVITIES:**

**SYSTEM STATE ON COMPLETION:** *The user has trained a new model, which he can now use for testing data.*

## 3.5 Testing data

**INITIAL ASSUMPTION:** *The user has an internet browser running and opened the system website, where he is not necessarily logged in as registered user. He has trained at least one model.*

**NORMAL:** *First of all the user has to go to the submenu item 'my models' where all his trained models are listed. From this list he has to pick the model he wants to test with by clicking on it. Now this model is marked and a new window opened with a list of his uploaded data sets. He picks one by clicking on it. This starts the testing process. After this process is completed the results of the tests are saved and notified to the user.*

**WHAT CAN GO WRONG:**

*The selected data is not compatible with the selected model. In this case*

testen und  
trainieren kann  
man nicht  
gleichzeitig? Bzw.  
eine Auswahl eines  
Teils der daten nur  
zum testen (wie  
beim trainieren) ist  
hier nicht möglich?

*an error message pops up and the user is able to choose another data set. The connection to the server is interrupted. In this case the user has to start the test again.*

**OTHER ACTIVITIES:**

**SYSTEM STATE ON COMPLETION:** *The user has got the result of a test.*