

Requirements Document for
SEDA
(System for Evaluation of Datamining
Algorithms)

Florian Hinkelmann Tobias Hinkelmann
Sebastian Ritzenhofen Felix Rauterberg
Dennis Maisenbacher

November 11, 2015

Glossary

arff Ist ein Datenformat, welches in ASCII-Text geschrieben ist und eine Liste von Instanzen beschreibt, die dieselben Attribute teilen.

(Quelle: <http://www.cs.waikato.ac.nz/ml/weka/arff.html> 11.11.2015). 21

ASCII Ist eine Zeichencodierung. (engl. American Standard Code for Information Interchange). 3

Bootstrap Ist ein System, welches HTML, Cascading Style Sheets und Javascript, jeweils auf Bildschirmgröße anpasst.. 20

Cascading Style Sheets Ist eine Sprache, mit der Gestaltungsanweisungen in HTML erstellt werden können.

(Quelle: de.wikipedia.org 11.11.2015). 3

csv Ist ein Datenformat, das den Aufbau einer Textdatei zur Speicherung oder zum Austausch einfach strukturierter Daten beschreibt. (engl. comma-separated values)

(Quelle: de.wikipedia.org 11.11.2015). 21

GUI Ist eine Grafische Benutzeroberfläche. (engl. Graphical User Interface). 13, 20, 24

HTML Ist eine Textbasierte Auszeichnungssprache zur Strukturierung digitaler Dokumente. (engl. Hypertext Markup Language)

(Quelle: de.wikipedia.org 11.11.2015). 3, 20

IEEE Ist ein weltweiter Berufsverband, der sich um die Förderung technologischer Innovationen zum Nutzen der Menschheit kümmert. (engl. Institute of Electrical and Electronics Engineers)

(Quelle: de.wikipedia.org 11.11.2015). 20

Java Ist eine Programmiersprache.. 20, 22, 24

Javascript Ist eine Sprache, mit der dynamisches HTML umgesetzt werden kann.

(Quelle: de.wikipedia.org 11.11.2015). 3

- JQuery** Ist eine freie Javascript Bibliothek.
(Quelle: de.wikipedia.org 11.11.2015). 20
- Open Source** Wird als Begriff für Software verwendet, deren Programmcode für jeden lizenzfrei zugänglich ist.
(Quelle: de.wikipedia.org 11.11.2015). 14
- Paket** Ist ein Ordner, der Trainingsdatensatz, Testdatensatz, Modell und Ergebnisdatensätze enthält.. 12, 13, 23
- RDF** Dies bezeichnet die technische Herangehensweise, um logische Aussagen zu treffen. (engl. Resource Description Framework)
(Quelle: de.wikipedia.org 11.11.2015). 19
- REST** Ist ein Programmierstil für Webanwendungen. (engl. Represational State Transfer)
(Quelle: de.wikipedia.org 11.11.2015). 11
- Salt** Ist eine zufällige gewählte Zeichenfolge, die zu einer Verschlüsselung zusätzliche Sicherheit bringt.
(Quelle: de.wikipedia.org 11.11.2015). 19
- SHA-256** Ist ein sicherer 256-Bit-Algorithmus zum verschlüsseln. (engl. Secure hash algorithm)
(Quelle: de.wikipedia.org 11.11.2015). 19
- SSL** Ist ein Verschlüsselungsprotokoll zur sicheren Datenübertragung. (engl. Secure Sockets Layer)
(Quelle: de.wikipedia.org 11.11.2015). 19
- Superadmin** Ist die Person, die das gesamte System überwacht und wartet. Diese Person besitzt alle Rechte des Systems.. 12, 21–24
- Webbrowser** Ist ein Programm zur Darstellung von Webseiten.. 17, 18
- WEKA** Ist ein Softwaretool, das verschiedene Techniken aus den Bereichen Maschinelles Lernen und Data-Mining bereitstellt. (engl. Waikato Environment for Knowledge Analysis)
(Quelle: de.wikipedia.org 11.11.2015). 20, 22

Contents

Glossary	3
I User Requirements	9
1 User Requirements	11
1.0.0.0.1 UR001	11
1.0.0.0.2 UR002	11
1.0.0.0.3 UR003	11
1.0.0.0.4 UR004	11
1.0.0.0.5 UR005	11
1.0.0.0.6 UR006	12
1.0.0.0.7 UR007	12
1.0.0.0.8 UR008	12
1.0.0.0.9 UR009	12
1.0.0.0.10 UR010	12
1.0.0.0.11 UR011	12
1.0.0.0.12 UR012	12
1.0.0.0.13 UR013	13
1.0.0.0.14 UR014	13
1.0.0.0.15 UR015	13
1.0.0.0.16 UR016	13
1.0.0.0.17 UR017	13
1.0.0.0.18 UR018	13
1.0.0.0.19 UR019	13
1.0.0.0.20 UR020	14
1.0.0.0.21 UR021	14
1.0.0.0.22 UR022	14
1.0.0.0.23 UR023	14
1.0.0.0.24 UR024	14
1.0.0.0.25 UR025	14
1.0.0.0.26 UR026	14

II	System Requirements	15
2	Non-Functional Requirements	17
2.1	Product Requirements	17
2.1.1	Usability Requirements	17
2.1.1.0.1	NFR001	17
2.1.1.0.2	NFR002	17
2.1.1.0.3	NFR003	17
2.1.1.0.4	NFR004	18
2.1.1.0.5	NFR005	18
2.1.2	Efficiency Requirements	18
2.1.2.1	Performance Requirements	18
2.1.2.1.1	NFR006	18
2.1.2.1.2	NFR007	18
2.1.2.1.3	NFR008	18
2.1.2.2	Space Requirements	18
2.1.2.2.1	NFR009	18
2.1.3	Dependability Requirements	19
2.1.4	Security Requirements	19
2.1.4.0.1	NFR010	19
2.1.4.0.2	NFR011	19
2.1.4.0.3	NFR012	19
2.1.4.0.4	NFR013	19
2.2	Organisational Requirements	19
2.2.1	Environmental Requirements	19
2.2.2	Operational Requirements	19
2.2.3	Development Requirements	19
2.2.3.0.1	NFR014	19
2.2.3.0.2	NFR015	20
2.2.3.0.3	NFR016	20
2.2.3.0.4	NFR017	20
2.3	External Requirements	20
2.3.1	Regulatory Requirements	20
2.3.2	Ethical Requirements	20
2.3.2.0.1	NFR014	20
2.3.3	Legislative Requirements	20
2.3.3.1	Accounting Requirements	20
2.3.3.2	Safety / Security Requirements	20
3	Functional Requirements	21
3.0.0.0.1	FR001	21
3.0.0.0.2	FR002	21
3.0.0.0.3	FR003	21
3.0.0.0.4	FR004	21
3.0.0.0.5	FR005	21
3.0.0.0.6	FR006	22

3.0.0.0.7	FR007	22
3.0.0.0.8	FR008	22
3.0.0.0.9	FR009	22
3.0.0.0.10	FR010	22
3.0.0.0.11	FR011	22
3.0.0.0.12	FR012	23
3.0.0.0.13	FR013	23
3.0.0.0.14	FR014	23
3.0.0.0.15	FR015	23
3.0.0.0.16	FR016	23
3.0.0.0.17	FR017	23
3.0.0.0.18	FR018	23
3.0.0.0.19	FR019	24
3.0.0.0.20	FR020	24
3.0.0.0.21	FR021	24
3.0.0.0.22	FR022	24
3.0.0.0.23	FR023	24
3.0.0.0.24	FR024	24
3.0.0.0.25	FR025	24
3.0.0.0.26	FR026	25
4	Scenarios	27
4.1	Registered user creates a new model	27
4.2	Registered user uses a model	28

Part I

User Requirements

Chapter 1

User Requirements

1.0.0.0.1 UR001

Statement *Das System soll eine Webanwendung zum Erstellen und Verwenden von Modellen, basierend auf Datamining- und Machinelearning- Algorithmen, sein.*

Priority *A*

1.0.0.0.2 UR002

Statement *Das System soll mit den Klassifikationsalgorithmen J48, Random Forest und SMO arbeiten können.*

Priority *A*

1.0.0.0.3 UR003

Statement *Das System soll REST-konform sein.*

Priority *A*

1.0.0.0.4 UR004

Statement *Das System soll alle Daten in einer Datenbank verwalten.*

Priority *A*

1.0.0.0.5 UR005

Statement *Die Datenbank soll eine Ordnerstruktur verwalten.*

Priority *A*

1.0.0.0.6 UR006

Statement *Das System soll durch einen Superadmin verwaltet und gewartet werden.*

Priority *A*

1.0.0.0.7 UR007

Statement *Eigene Algorithmen sollen als Plugins hochgeladen und nach Überprüfung des Superadmins freigeschaltet werden können.*

Priority *A*

1.0.0.0.8 UR008

Statement *Durch Anwendung der Trainingsdaten auf einen vom Benutzer gewählten Algorithmus soll ein Modell erstellt werden.*

Priority *A*

1.0.0.0.9 UR009

Statement *Bereits bestehende Modelle sollen nur von angemeldeten Benutzern hochgeladen werden können.*

Priority *A*

1.0.0.0.10 UR010

Statement *Das System soll als angemeldeter Benutzer und anonym verwendet werden können.*

Priority *A*

1.0.0.0.11 UR011

Statement *Bei der Anmeldung eines Benutzers soll automatisch ein neues, leeres Paket für ihn erstellt werden, auf welches anfangs nur er selbst Lese- und Schreibrechte besitzt.*

Priority *A*

1.0.0.0.12 UR012

Statement *Der anonyme Benutzer soll mehrere Einschränkungen gegenüber des angemeldeten haben.*

Priority *A*

1.0.0.0.13 UR013

Statement *Pakete sollen von allen Benutzern erstellt werden können.*

Priority *A*

1.0.0.0.14 UR014

Statement *Der angemeldete Benutzer darf für seine Pakete die Lese- und Schreibrechte für andere Nutzer auswählen. Außerdem darf er weitere Admins für dieses Paket benennen.*

Priority *A*

1.0.0.0.15 UR015

Statement *Der Benutzer soll auf bestehende Modelle Trainingsdaten anwenden können, sofern er die Berechtigung dazu hat.*

Priority *A*

1.0.0.0.16 UR016

Statement *Der Nutzer kann sich die Ergebnisse, Modelle und Datensätze des jeweiligen Pakets oder das Paket selbst herunterladen oder direkt in der GUI anzeigen lassen, sofern er die Rechte dazu hat.*

Priority *A*

1.0.0.0.17 UR017

Statement *Die Resultate sollen sowohl in Textform als auch grafisch zur Verfügung gestellt werden.*

Priority *A*

1.0.0.0.18 UR018

Statement *Der Benutzer soll auswählen können, welche Ergebnisparameter als Resultate ausgegeben werden sollen.*

Priority *A*

1.0.0.0.19 UR019

Statement *Der Nutzer soll die laufende Berechnung des Modells abbrechen können.*

Priority *A*

1.0.0.0.20 UR020

Statement *Das System soll Open Source sein.*

Priority *A*

1.0.0.0.21 UR021

Statement *Das System soll von allen gängigen Browsern, unter anderem auch auf mobilen Geräten, erreichbar sein.*

Priority *A*

1.0.0.0.22 UR022

Statement *Das System soll eine Vorhersage in kurzer Zeit ausgeben können.*

Priority *A*

1.0.0.0.23 UR023

Statement *Die Benutzeroberfläche soll in Englisch gehalten werden.*

Priority *A*

1.0.0.0.24 UR024

Statement *Das System soll einfach bedienbar und schnell erlernbar sein.*

Priority *A*

1.0.0.0.25 UR025

Statement *Auf Modelle sollen Lizenzen vergeben werden können.*

Priority *B*

1.0.0.0.26 UR026

Statement *Bestehende Modelle sollen nicht verändert werden können.*

Priority *A*

Part II

System Requirements

Chapter 2

Non-Functional Requirements

2.1 Product Requirements

2.1.1 Usability Requirements

2.1.1.0.1 NFR001

Statement *Der Umgang mit dem System soll für den Nutzer innerhalb von 10 Stunden erlernbar sein. (Siehe UR024)*

Priority *A*

2.1.1.0.2 NFR002

Statement *Die Administration des Systems soll innerhalb von 70 Stunden erlernbar sein. (Siehe UR024)*

Priority *A*

2.1.1.0.3 NFR003

Statement *Das System soll auf den Webbrowsern*

- *Mozilla Firefox ab Version 32.0*
- *Opera ab Version 27.0*
- *Safari ab Version 8.0*
- *Google Chrome ab Version 38.0.2125*
- *Microsoft Edge ab Version 1.0*
- *Internet Explorer ab Version 11*

verfügbar sein. (Siehe UR021)

Priority *A*

2.1.1.0.4 NFR004

Statement *Das System soll auf den gängigen mobilen Webbrowsern*

- *Android Browser*
- *Chrome*
- *Safari*
- *Internet Explorer Mobile*

verfügbar sein. (Siehe UR021)

Priority *B*

2.1.1.0.5 NFR005

Statement *Es soll keine Desktop-Applikation geben. (Siehe UR001)*

Priority *A*

2.1.2 Efficiency Requirements**2.1.2.1 Performance Requirements****2.1.2.1.1 NFR006**

Statement *Die Zeit, in der ein Modell errechnet wird, ist irrelevant.*

Priority *A*

2.1.2.1.2 NFR007

Statement *Das System soll auf einem Modell eine Vorhersage binnen 5 Sekunden treffen. (Siehe UR022)*

Priority *A*

2.1.2.1.3 NFR008

Statement *Das System soll auf einem Desktop PC laufen können, der mindestens 32GB RAM und eine CPU mit 8 Kernen besitzt.*

Priority *A*

2.1.2.2 Space Requirements**2.1.2.2.1 NFR009**

Statement *Es soll keine Beschränkung für den benötigten Speicherplatz der Datenbank geben. (Siehe UR004)*

Priority *A*

2.1.3 Dependability Requirements

2.1.4 Security Requirements

2.1.4.0.1 NFR010

Statement *Alle Nutzerdaten sollen sicher sein und werden nicht außerhalb des Systems verwendet.*

Priority *A*

2.1.4.0.2 NFR011

Statement *Der Nutzer soll nach 15 Minuten Inaktivität automatisch ausgeloggt werden.*

Priority *A*

2.1.4.0.3 NFR012

Statement *Die Kommunikation zwischen Server und Client soll mit SSL verschlüsselt werden. (Siehe UR001)*

Priority *A*

2.1.4.0.4 NFR013

Statement *Passwörter der Benutzer sollen mit SHA-256 und einem Salt sicher verschlüsselt sein. (Siehe UR010)*

Priority *A*

2.2 Organisational Requirements

2.2.1 Environmental Requirements

2.2.2 Operational Requirements

2.2.3 Development Requirements

2.2.3.0.1 NFR014

Statement *Das System soll mit Datenbanken im RDF-Format arbeiten. (Siehe UR004)*

Priority *A*

2.2.3.0.2 NFR015

Statement *Die Basis-Algorithmen sollen von WEKA zur Verfügung gestellt werden. (Siehe UR002)*

Priority *A*

2.2.3.0.3 NFR016

Statement *Das System soll in Java programmiert sein.*

Priority *A*

2.2.3.0.4 NFR017

Statement *Die GUI soll HTML 5, Bootstrap und JQuery konform sein. (Siehe UR016)*

Priority *A*

2.3 External Requirements**2.3.1 Regulatory Requirements****2.3.2 Ethical Requirements****2.3.2.0.1 NFR014**

Statement *Das System muss dem IEEE Standard für ethische Software genügen.*

Priority *A*

2.3.3 Legislative Requirements**2.3.3.1 Accounting Requirements****2.3.3.2 Safety / Security Requirements**

Chapter 3

Functional Requirements

3.0.0.0.1 FR001

Statement *Trainings- und Testdaten sollen nur in „csv“ und „arff“ Format hochgeladen werden können. (Siehe UR008)*

Priority *A*

3.0.0.0.2 FR002

Statement *Das System soll die übergebenen Datensätze auf dem Server verarbeiten. (Siehe UR001)*

Priority *A*

3.0.0.0.3 FR003

Statement *Anonyme Benutzer sollen über eine Laufzeitbeschränkung reguliert sein, welche der Superadmin einstellen kann. (Siehe UR012)*

Priority *A*

3.0.0.0.4 FR004

Statement *Anonyme Benutzer sollen ein Upload-Limit haben, über welches der Superadmin verfügen kann. (Siehe UR012)*

Priority *A*

3.0.0.0.5 FR005

Statement *Die Ausgabe soll entsprechend des verwendeten Algorithmus dargestellt werden. (Siehe UR007)*

Priority *A*

3.0.0.0.6 FR006

Statement *Bei Registrierung sollen E-Mail-Adresse, Passwort und optional ein Name angegeben werden. (Siehe UR010)*

Priority *A*

3.0.0.0.7 FR007

Statement *Es sollen standardmäßig zwei Modelle gleichzeitig berechnet werden können. Alle weiteren Berechnungen sind in einer Warteschlange. Der Superadmin soll einstellen können, wie viele gleichzeitig berechnet werden können. (Siehe UR008)*

Priority *A*

3.0.0.0.8 FR008

Statement *Anhand der Parameter der Trainingsdaten sollen dem Benutzer passende Algorithmen zur Auswahl gestellt werden. (Siehe UR002)*

Priority *A*

3.0.0.0.9 FR009

Statement *Anhand der Parameter der Testdaten soll geprüft werden, ob sie auf das gewählte Modell anwendbar sind. (Siehe UR008)*

Priority *A*

3.0.0.0.10 FR010

Statement *Beim Hochladen eines Datensatzes soll der Benutzer zwei Möglichkeiten haben:*

- *Der komplette Datensatz soll aus Trainingsdaten bestehen. Die Testdaten sollen in diesem Fall im Nachhinein hochgeladen werden.*
- *Der Datensatz wird nach Wahl des Benutzers in Trainings- und Testdaten unterteilt. Im Nachhinein können weitere Testdaten hochgeladen werden.*

(Siehe UR008)

Priority *A*

3.0.0.0.11 FR011

Statement *Hochgeladene Algorithmen sollen als .jar gepackt sein und von der (Java-)WEKA-Klasse Classifier erben. (Siehe UR007)*

Priority *A*

3.0.0.0.12 FR012

Statement *Wenn der Speicher des Servers voll ist, wird eine Fehlermeldung geworfen. (Siehe UR001)*

Priority *A*

3.0.0.0.13 FR013

Statement *Zu jedem Paket sollen Zugriffsrechte vergeben und gespeichert werden. (Siehe UR003)*

Priority *A*

3.0.0.0.14 FR014

Statement *Werden Trainingsdaten für einen Algorithmus eingegeben, für die bereits ein Modell mit genau diesen Trainingsdaten existiert, soll der Nutzer entscheiden können, ob das vorhandene Modell genutzt oder ein neues Modell angelegt wird. (Siehe UR026)*

Priority *A*

3.0.0.0.15 FR015

Statement *Alles in einem Paket sowie das Paket selbst soll eine ID besitzen. (Siehe UR005)*

Priority *A*

3.0.0.0.16 FR016

Statement *Ein Ordner soll Paket heißen und Trainingsdatensatz, Testdatensatz, Modell und Ergebnisdatensätze enthalten. Pakete können geschachtelt werden. (Siehe UR005)*

Priority *A*

3.0.0.0.17 FR017

Statement *Das System soll auf verschiedene Server zugreifen können. (Siehe UR003)*

Priority *A*

3.0.0.0.18 FR018

Statement *Der Superadmin soll über alle Inhalte des Systems verfügen und die Eigenschaften konfigurieren können. (Siehe UR006)*

Priority *A*

3.0.0.0.19 FR019

Statement *Der Superadmin soll eine erweiterte GUI erhalten, in der seine erweiterten Funktionen ausführbar sind. (Siehe UR006)*

Priority *A*

3.0.0.0.20 FR020

Statement *Algorithmen sollen von angemeldeten Benutzern der System-Bibliothek hinzugefügt werden können. (Siehe UR007)*

Priority *A*

3.0.0.0.21 FR021

Statement *Modelle sollen als serialisiertes Java Objekt portierbar sein. (Siehe UR009)*

Priority *A*

3.0.0.0.22 FR022

Statement *Standartmäßig sollen Modelle von nicht angemeldeten Nutzern nach 30 Tagen gelöscht werden. Der Superadmin kann über diesen Zeitraum nachträglich bestimmen. Modelle von angemeldeten Benutzern bleiben auf unbestimmte Zeit erhalten. (Siehe UR012)*

Priority *A*

3.0.0.0.23 FR023

Statement *Anonyme Benutzer dürfen weder Lese- noch Schreibrechte festlegen. (Siehe UR012)*

Priority *A*

3.0.0.0.24 FR024

Statement *Extern hochgeladene Modelle sollen vom Superadmin erst freigegeben werden müssen. (Siehe UR009)*

Priority *A*

3.0.0.0.25 FR025

Statement *Modelle von anonymen Nutzern sollen für alle nutzbar sein. (Siehe UR012)*

Priority *A*

3.0.0.0.26 FR026

Statement *Ein angemeldeter Benutzer soll die Lese- und Schreibrechte für alle Benutzer, bestimmte Benutzergruppen oder einzelne Benutzer festlegen können. (Siehe UR014)*

Priority *A*

Chapter 4

Scenarios

4.1 Registered user creates a new model

INITIAL ASSUMPTION: *Der Benutzer hat in seinem Webbrowser den Web-service geöffnet. Er ist mit seinem Benutzerkonto angemeldet.*

NORMAL: *Der Nutzer erstellt zunächst ein Paket oder wählt ein bereits vorhandenes Paket, auf das er Zugriff hat, aus. Falls er Adminrechte auf das Paket hat, kann er die Lese- und Schreibrechte festlegen. Nun lädt er einen Datensatz im csv oder arff Format in das Paket hoch. Diesen kann der Nutzer nach Belieben in Trainings- und Testdaten unterteilen. Es wird ihm eine Vorauswahl an Algorithmen angeboten, die mit dem Datensatz kompatibel sind. Nach der Auswahl des Algorithmus errechnet der Server das Modell. Das fertig errechnete Modell wird im selben Paket gespeichert. Falls bereits Testdaten in dem anfangs hochgeladenen Datensatz enthalten waren, wird das Ergebnis der Vorhersage ebenfalls im Paket gespeichert.*

WHAT CAN GO WRONG:

- *Der Nutzer verliert die Internetverbindung.*
- *Das Modell besteht schon. Der Nutzer hat die Wahl, das vorhandene Modell zu benutzen oder ein neues Modell zu berechnen.*
- *Der Server ist bereits ausgelastet. In diesem Fall wird die Berechnung des Modells hinten angestellt.*
- *Der Speicher des Servers ist voll. Es wird eine Fehlermeldung angezeigt.*

OTHER ACTIVITIES:

SYSTEM STATE ON COMPLETION: *Das Modell ist berechnet, es können weitere Testdaten darauf angewendet werden.*

4.2 Registered user uses a model

INITIAL ASSUMPTION: *Der Benutzer hat in seinem Webbrowser den Web-service geöffnet. Er ist mit seinem Benutzerkonto angemeldet und befindet sich in dem entsprechenden Paket.*

NORMAL: *Der Nutzer kann Testdaten im csv oder arff Format in das Paket hochladen. Es können nun die auszuwertenden Parameter, die in der Ausgabe erscheinen sollen, ausgewählt werden. Die vom Server errechneten Vorhersagen für das Modell, basierend auf den Trainingsdaten, werden dann im Paket gespeichert.*

WHAT CAN GO WRONG:

- *Die Parameter des Datensatzes sind nicht kompatibel zu dem Modell. Es gibt eine Fehlermeldung.*
- *Der Speicher des Servers ist voll. Es wird eine Fehlermeldung angezeigt.*
- *Der Nutzer verliert die Internetverbindung.*

OTHER ACTIVITIES:

SYSTEM STATE ON COMPLETION: *Die Vorhersagen sind bereit, angezeigt oder heruntergeladen zu werden.*