

Requirements Document for
stupid data-mining project
v1.0 Ant

Felix Brand

Joshua Witte
Lukas Weller

Simon Shamon

November 11, 2015

Contents

I	User Requirements	7
	0.0.0.0.1	9
	0.0.0.0.2	9
	0.0.0.0.3	9
	0.0.0.0.4	9
	0.0.0.0.5	9
	0.0.0.0.6	9
	0.0.0.0.7	9
	0.0.0.0.8	9
	0.0.0.0.9	10
	0.0.0.0.10	10
	0.0.0.0.11	10
	0.0.0.0.12	10
	0.0.0.0.13	10
	0.0.0.0.14	10
	0.0.0.0.15	10
	0.0.0.0.16	10
	0.0.0.0.17	11
	0.0.0.0.18	11
	0.0.0.0.19	11
	0.0.0.0.20	11
	0.0.0.0.21	11
	0.0.0.0.22	11
	0.0.0.0.23	11
	0.0.0.0.24	11
	0.0.0.0.25	12
	0.0.0.0.26	12
	0.0.0.0.27	12
	0.0.0.0.28	12
	0.0.0.0.29	12
	0.0.0.0.30	12
	0.0.0.0.31	12
	0.0.0.0.32	13
	0.0.0.0.33	13
	0.0.0.0.34	13

0.0.0.0.35	13
0.0.0.0.36	13
0.0.0.0.37	13
0.0.0.0.38	13
0.0.0.0.39	13
0.0.0.0.40	14

II System Requirements 15

1 Nicht-Funktionale Anforderungen 17

1.1 Produkt Anforderungen	17
1.1.1 Usability Requirements	17
1.1.1.0.41	17
1.1.1.0.42	17
1.1.1.0.43	17
1.1.1.0.44	17
1.1.1.0.45	18
1.1.1.0.46	18
1.1.2 Efficiency Requirements	18
1.1.2.1 Performance Requirements	18
1.1.2.1.1	18
1.1.2.1.2	18
1.1.2.1.3	18
1.1.2.1.4	18
1.1.2.2 Space Requirements	18
1.1.2.2.1	18
1.1.2.2.2	19
1.1.2.2.3	19
1.1.3 Dependability Requirements	19
1.1.3.0.4	19
1.1.3.0.5	19
1.1.4 Security Requirements	19
1.1.4.0.6	19
1.1.4.0.7	19
1.2 Organisationsvorschriften	20
1.2.1 Umgebungsanforderungen	20
1.2.2 Betriebsanforderungen	20
1.2.3 Entwicklungsanforderungen	20
1.2.3.0.8	20
1.3 Äußere Anforderungen	20
1.3.1 Regulatory Requirements	20
1.3.1.0.9 NFR024	20
1.3.2 Ethical Requirements	20
1.3.2.0.10 NFR025	20
1.3.3 Rechtliche Anforderungen	20

<i>CONTENTS</i>	5
1.3.3.1 Accounting Requirements	20
1.3.3.2 Safety / Security Requirements	20
1.3.3.2.1 NFR026	20
2 Functional Requirements	21
2.0.3.2.2	21
2.0.3.2.3	21
2.0.3.2.4	21
3 Scenarios	23
3.1 Im System anmelden	23
3.2 Datensätze hochladen	23
3.3 Ein Modell erstellen	24

Part I

User Requirements

0.0.0.0.1

Aussage *Es wird ein Online-Framework für Datamining und Machinelearning entwickelt.*

Priorität *A+*

0.0.0.0.2

Aussage *Benutzer können sich registrieren oder das Programm anonym benutzen.*

Priorität *A*

0.0.0.0.3

Aussage *Anonyme Nutzer haben eingeschränkte Rechte.*

Priorität *A*

0.0.0.0.4

Aussage *Es gibt mindestens einen Administrator mit erweiterten Rechten.*

Priorität *A*

0.0.0.0.5

Aussage *Angemeldete Nutzer können Gruppen bilden um gemeinsam an einem Datensatz/Paket zu arbeiten.*

Priorität *A*

0.0.0.0.6

Aussage *Die Benutzer sind lokal auf dem Server gespeichert, auf einem anderen Server muss man sich neu anmelden.*

Priorität *B*

0.0.0.0.7

Aussage *Das Programm arbeitet mit in Java implementierten Algorithmen.*

Priorität *A*

0.0.0.0.8

Aussage *Die Algorithmen dienen der Klassifikation und evtl. Clustering.*

Priorität *A*

0.0.0.0.9

Aussage *Die Algorithmen SMO, J48 und Random Forest sind bereits implementiert.*

Priorität *A*

0.0.0.0.10

Aussage *Laufende Algorithmen können abgebrochen werden.*

Priorität *A*

0.0.0.0.11

Aussage *Die Algorithmen werden für den vorliegenden Datensatz vorsortiert, nur anwendbare werden angezeigt.*

Priorität *A*

0.0.0.0.12

Aussage *Nutzer haben die Möglichkeit eigene Algorithmen als .jar Datei zum Programm hinzu zu fügen.*

Priorität *A*

0.0.0.0.13

Aussage *Es steht ein festes Interface für die Algorithmen zur Verfügung.*

Priorität *A*

0.0.0.0.14

Aussage *Algorithmen werden vom Admin geprüft, bevor sie bereitgestellt werden.*

Priorität *B*

0.0.0.0.15

Aussage *Das Programm steht über eine Web-Oberfläche zur Verfügung. Eine gesonderte mobile Website ist nicht gefordert.*

Priorität *A*

0.0.0.0.16

Aussage *Die Website steht in Englischer Sprache zur Verfügung.*

Priorität *A*

0.0.0.0.17

Aussage *Die Website wird in HTML programmiert.*

Priorität *A*

0.0.0.0.18

Aussage *Dem Nutzer steht eine Übersicht über alle Pakete zur Verfügung, auf die er Zugriff hat.*

Priorität *A*

0.0.0.0.19

Aussage *Die Modelle werden in einer RDF-Datenbank gespeichert.*

Priorität *A*

0.0.0.0.20

Aussage *Die Datenbank unterstützt SPARQL*

Priorität *A*

0.0.0.0.21

Aussage *Der Zugriff auf das System ist sicher. Niemand kann Daten für die er kein Zugriffsrecht besitzt manipulieren.*

Priorität *A*

0.0.0.0.22

Aussage *Mithilfe der Algorithmen sollen aus den Daten Modelle erstellt und gespeichert werden können (in RDF-Format).*

Priorität *A*

0.0.0.0.23

Aussage *Vorhandene Modelle sollen auf andere Daten angewendet werden können und dadurch Vorhersagen machen oder die Güte des Modells bewerten.*

Priorität *A*

0.0.0.0.24

Aussage *Alle Daten sollen auf einem Server gespeichert werden.*

Priorität *A*

0.0.0.0.25

Aussage *Die Daten sollen von den Nutzern in die Datenbank geladen werden können.(in arff bzw. csv Format*

Priorität *A*

0.0.0.0.26

Aussage *Der Nutzer kann separate Datenbestände zum Lernen und Testen hochladen, oder einen Datenbestand der auf dem Server in 2 Teile geteilt wird(Verhältnis vom Nutzer festzulegen).*

Priorität *A*

0.0.0.0.27

Aussage *Daten anonymer Nutzer werden nach einer vom Admin festgelegten Frist vom Server gelöscht.*

Priorität *A*

0.0.0.0.28

Aussage *Die Daten werden kategorisiert gespeichert.*

Priorität *A*

0.0.0.0.29

Aussage *Daten anonymer Nutzer werden in einem gemeinsamen Anonym-Paket gespeichert.*

Priorität *A*

0.0.0.0.30

Aussage *Einmal erstellte Modelle sollen nicht mehr verändert werden können.*

Priorität *A*

0.0.0.0.31

Aussage *Es sollen auch Modelle von Nutzern hochgeladen werden können.*

Priorität *B*

0.0.0.0.32

Aussage *Wird Versucht ein Modell zu erstellen, dass bereits vorhanden ist, soll gefragt werden, ob stattdessen das bereits berechnete Modell aufgerufen werden soll.*

Priorität *B*

0.0.0.0.33

Aussage *Das System soll Pakete anbieten. Jedes Paket soll Datensätze, Modelle und Ergebnisse enthalten.*

Priorität *A*

0.0.0.0.34

Aussage *Jeder Benutzer soll Pakete erstellen können.*

Priorität *A*

0.0.0.0.35

Aussage *Jedes Paket ist über seine ID such- und adressierbar*

Priorität *A*

0.0.0.0.36

Aussage *Die Pakete sollen Zugriffsrechte haben. Standard ist Privat.*

Priorität *A*

0.0.0.0.37

Aussage *Der Benutzer soll die Zugriffsrechte des Pakets ändern können. z.B. einer anderen Gruppe Schreibrechte auf ein bestimmtes Paket geben.*

Priorität *A*

0.0.0.0.38

Aussage *Der Benutzer kann Pakete öffentlich machen.*

Priorität *A*

0.0.0.0.39

Aussage *Nach dem Bearbeiten der Datensätze soll der Kunde die Ergebnisse kriegen. Die Ergebnisse sollen selektier- und formatierbar sein.*

Priorität *A*

0.0.0.0.40

Aussage *Der Kunde soll die Ergebnisse herunterladen können. Der Kunde soll wählen können, ob er die Ergebnisse in HTML oder PDF herunterladen möchte.*

Priorität *A*

Part II

System Requirements

Chapter 1

Nicht-Funktionale Anforderungen

1.1 Produkt Anforderungen

1.1.1 Usability Requirements

1.1.1.0.41

Statement *Eine Internetverbindung soll ausreichen, um das System zu erreichen und zu benutzen. Firewall- oder Routerbeschränkungen sollen nicht existieren.*

Priority *A*

1.1.1.0.42

Statement *Die Web-Anwendung soll auf die folgenden Internetbrowser nutzbar sein(inklusive neuerer Versionen): Internet explorer 9, Firefox 30, Opera 12, Chrome 45.*

Priority *A*

1.1.1.0.43

Statement *Die Sprache des Systems ist Englisch.*

Priority *A*

1.1.1.0.44

Statement *Es soll KEINE Desktop-, IOS- oder Android-Anwendung für das System geben.*

Priority *A*

1.1.1.0.45

Statement *Alle grundlegenden Funktionen des Systems sollen für einen Benutzer, der ein Bachelor of Science in Informatik absolviert hat, in 10 Stunden erlernbar sein.*

Priority *A*

1.1.1.0.46

Statement *Das System soll leicht sein zu verwalten und zu steuern. Der Admin, der sich schon mit der Systemverwaltung auskennt, soll alle erforderlichen Funktionen in 70 Stunden erlernen können.*

Priority *A*

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

Statement *Das Programm wird auf Computern mit mindestens 32 GB Ram, 8 Prozessorernen ausgeführt.*

Priority *A*

1.1.2.1.2

Statement *Die Rechenzeit registrierter Nutzer ist nicht beschränkt.*

Priority *A*

1.1.2.1.3

Statement *Jeder 999. User darf alles.*

Priority *A*

1.1.2.1.4

Statement *Die Rechenzeit anonymer nutzer darf einen vom Admin festgelegten Grenzwert nicht überschreiben.*

Priority *A*

1.1.2.2 Space Requirements**1.1.2.2.1**

Statement *Registrierte Nutzer können beliebig große Datenhochladen.*

Priority *A*

1.1.2.2.2

Statement *Uploads anonymer Nutzer dürfen einen vom Admin festgelegten Grenzwert nicht überschreiten.*

Priority *A*

1.1.2.2.3

Statement *Daten anonymer Nutzer werden nach einer vom Admin festgelegten Zeit ohne Aufruf vom Server gelöscht,*

Priority *A*

1.1.3 Dependability Requirements

1.1.3.0.4

Statement *Falls das System abstürzen sollte, wird automatisch versucht es erneut zu starten und der Administrator bekommt eine Meldung über den Systemabbruch.*

Priority *A*

1.1.3.0.5

Statement *Das System sollte rund um die Uhr erreichbar sein.*

Priority *A*

1.1.4 Security Requirements

1.1.4.0.6

Statement *Passwörter werden nicht im Klartext gespeichert.*

Priority *A*

1.1.4.0.7

Statement *Daten können nur von Nutzern aufgerufen/verändert werden, die dazu berechtigt sind.*

Priority *A*

1.2 Organisationsvorschriften

1.2.1 Umgebungsanforderungen

1.2.2 Betriebsanforderungen

1.2.3 Entwicklungsanforderungen

1.2.3.0.8

Statement *Die Software soll in gültigem HTML5 und Java geschrieben sein.*

Priority *A*

1.3 Äußere Anforderungen

1.3.1 Regulatory Requirements

1.3.1.0.9 NFR024

Statement *Das Programm wird unter GPL Version 3 veröffentlicht.*

Priority *A*

1.3.2 Ethical Requirements

1.3.2.0.10 NFR025

Statement *die "IEEE Code of Ethics" werden beachtet.*

Priority *A*

1.3.3 Rechtliche Anforderungen

1.3.3.1 Accounting Requirements

1.3.3.2 Safety / Security Requirements

1.3.3.2.1 NFR026

Statement *Das Programm verstößt nicht gegen geltendes deutsches Recht.*

Priority *A*

Chapter 2

Functional Requirements

2.0.3.2.2

Statement *Nutzer können sich mit Name, Email-Adressel und Passwort registrieren.*

Priority *A*

2.0.3.2.3

Statement *Die Email-Adresse muss bestätigt werden.*

Priority *B*

2.0.3.2.4

Statement *Anonyme Nutzer können keinen Gruppen beitreten.*

Priority *A*

Chapter 3

Scenarios

3.1 Im System anmelden

URSPRÜNGLICHE ANNAHME: *Der Benutzer öffnet die Webseite in seinem Browser.*

NORMAL: *Der Benutzer gibt in die zugehörigen Textfelder seine e-mail-Adresse und Passwort ein und klickt auf den login-button oder der Benutzer klickt auf den login-as-guest-button*

WAS KANN SCHIEFGEHEN: *Die Verbindung zum Server kann nicht hergestellt werden und es wird eine Fehlermeldung angezeigt.
Der Nutzer gibt eine unzulässige Kombination von e-mail und Passwort an. Es wird eine Meldung angezeigt, dass der Benutzername oder das Passwort falsch ist und es wird gefragt, ob der Nutzer sich stattdessen als Gast anmelden will.
Die Verbindung mit dem Internet ist unterbrochen.*

ANDERE AKTIVITÄTEN:

SYSTEMZUSTAND BEI ERFOLG: *Der Nutzer ist eingeloggt und kann das System benutzen.*

3.2 Datensätze hochladen

URSPRÜNGLICHE ANNAHME: *Der Benutzer öffnet die Seite in seinem Browser.*

NORMAL: *Der Benutzer klickt auf "Package" oben links. Dann wählt er ein Paket aus, oder erstellt ein neues Paket. Er kann danach Datensätze hochladen. Dafür klickt er auf den Button "Dateien hochladen". Dann erscheint ein Fenster, wo er die Dateien aus seinem lokalen Computer*

auswählen kann, und klickt er auf OK. Die Dateien werden danach hochgeladen.

WAS KANN SCHIEFGEHEN: *Der Benutzer versucht eine Datei hochzuladen, dessen Typ im System nicht erlaubt ist, oder der verfügbare Speicherplatz wird überschritten. In beiden Fällen bekommt er eine Fehlermeldung und der Prozess wird abgebrochen. Die Verbindung zum Server wird unterbrochen. Es wird eine dementsprechende Fehlermeldung gezeigt.*

ANDERE AKTIVITÄTEN:

SYSTEMZUSTAND BEI ERFOLG: *Der Nutzer hat neue Daten in die Datenbank geladen und freut sich.*

3.3 Ein Modell erstellen

URSPRÜNGLICHE ANNAHME: *Der Nutzer hat sich auf der Webseite angemeldet.*

NORMAL: *Der Nutzer klickt auf den train-a-new-model-button und ein Fenster öffnet sich wo der Nutzer einen Datensatz aus der Datenbank auswählt, mit dem er das Modell erstellen will. Dann wählt er in einem weiteren Menü den zu nutzenden Algorithmus und setzt die notwendigen Parameter. Sobald der train-now-button gedrückt wird, startet die Berechnung des Modells und lizenzfreie Fahrstuhlmusik wird abgespielt bis die Berechnung beendet ist. Danach erscheint eine Meldung, dass die Berechnung erfolgreich abgeschlossen wurde, mit den Optionen das Modell direkt auf andere Daten anzuwenden, zu verwerfen oder abzuspeichern.*

WAS KANN SCHIEFGEHEN: *Es kann nicht auf den Datensatz oder das Modell zugegriffen werden, woraufhin eine Fehlermeldung erscheint. Der Benutzer gibt zu wenige und/oder unzulässige Parameter an, woraufhin eine Meldung erscheint, dass die Parameter für das gewählte Modell nicht zulässig sind. Es werden zu viele andere Anfragen an den Server gestellt. Es erscheint ein Warnhinweis, dass der Vorgang zur Zeit nicht ausgeführt werden kann, da der Server ausgelastet ist. Zusätzlich wird die Position in der Warteschlange der ausstehenden Anfragen angezeigt.*

ANDERE AKTIVITÄTEN:

SYSTEMZUSTAND BEI ERFOLG: *Der Nutzer hat ein Modell erstellt, und die Option dieses in der Datenbank als Paket mit den zugehörigen Daten zu speichern.*