

Master Thesis Kick-Off: Interactive Analysis of a Corpus of General Terms and Conditions for Variability Modeling

David Koller 23.11.2018

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
wwwmatthes.in.tum.de

Key Facts



Title: Interactive Analysis of a Corpus of General Terms and Conditions for Variability Modeling

Author: B.Sc. David Koller

Advisor: M.Sc. Ingo Glaser

Supervisor: Prof. Dr. Florian Matthes

Start: October 15th, 2018

End: April 15th, 2019

Introduction

- Motivation
- Problem Statement

Research Questions

- Interactive Machine Learning Introduction
- Data Collection and Preprocessing
- Text Classification

Approach

- Workflow
- Timetable

Problem:

- No standardized Terms & Conditions
- Question of legality of certain clauses

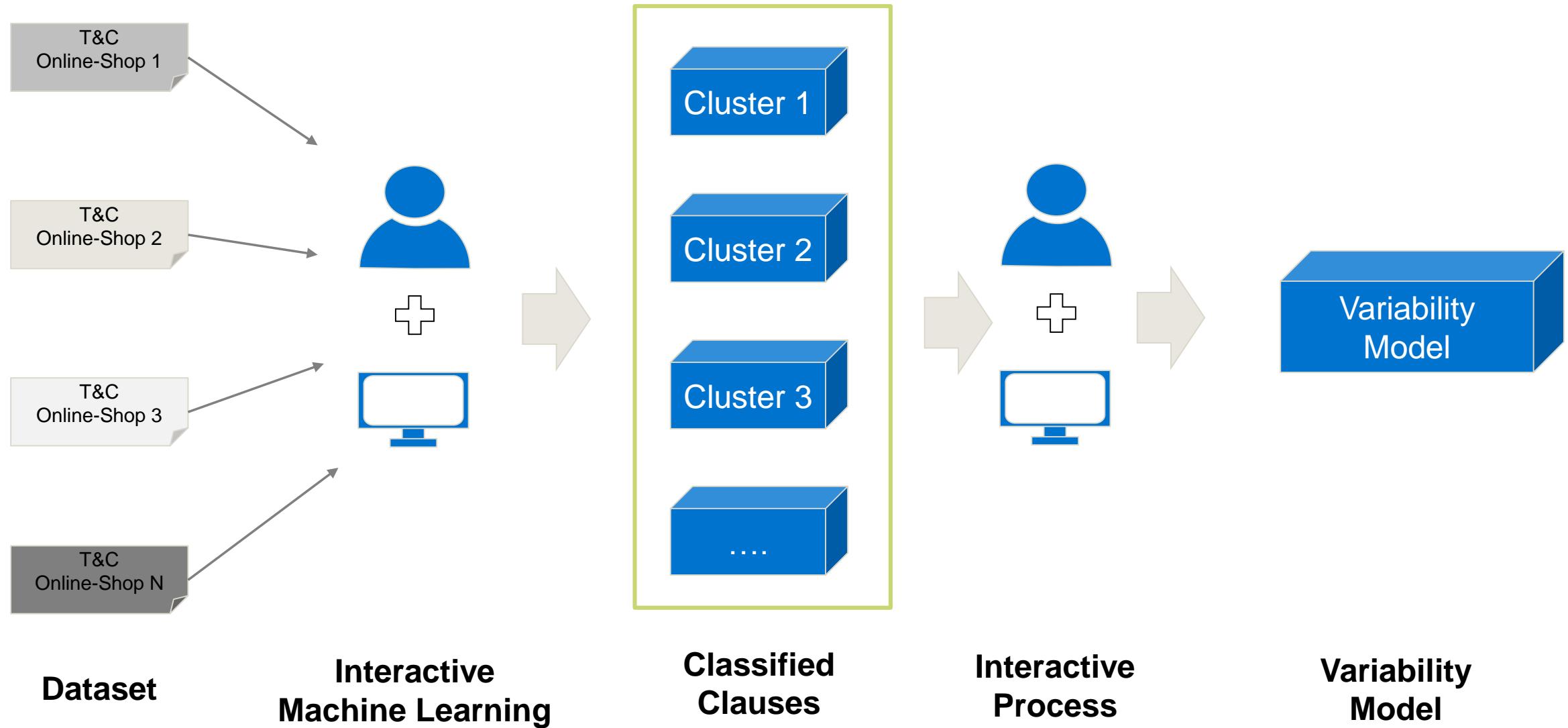
**Solution:**

- Classifying clauses from Terms & Conditions
- Identifying universally needed clauses and dependencies
- Recognizing unlawful clauses

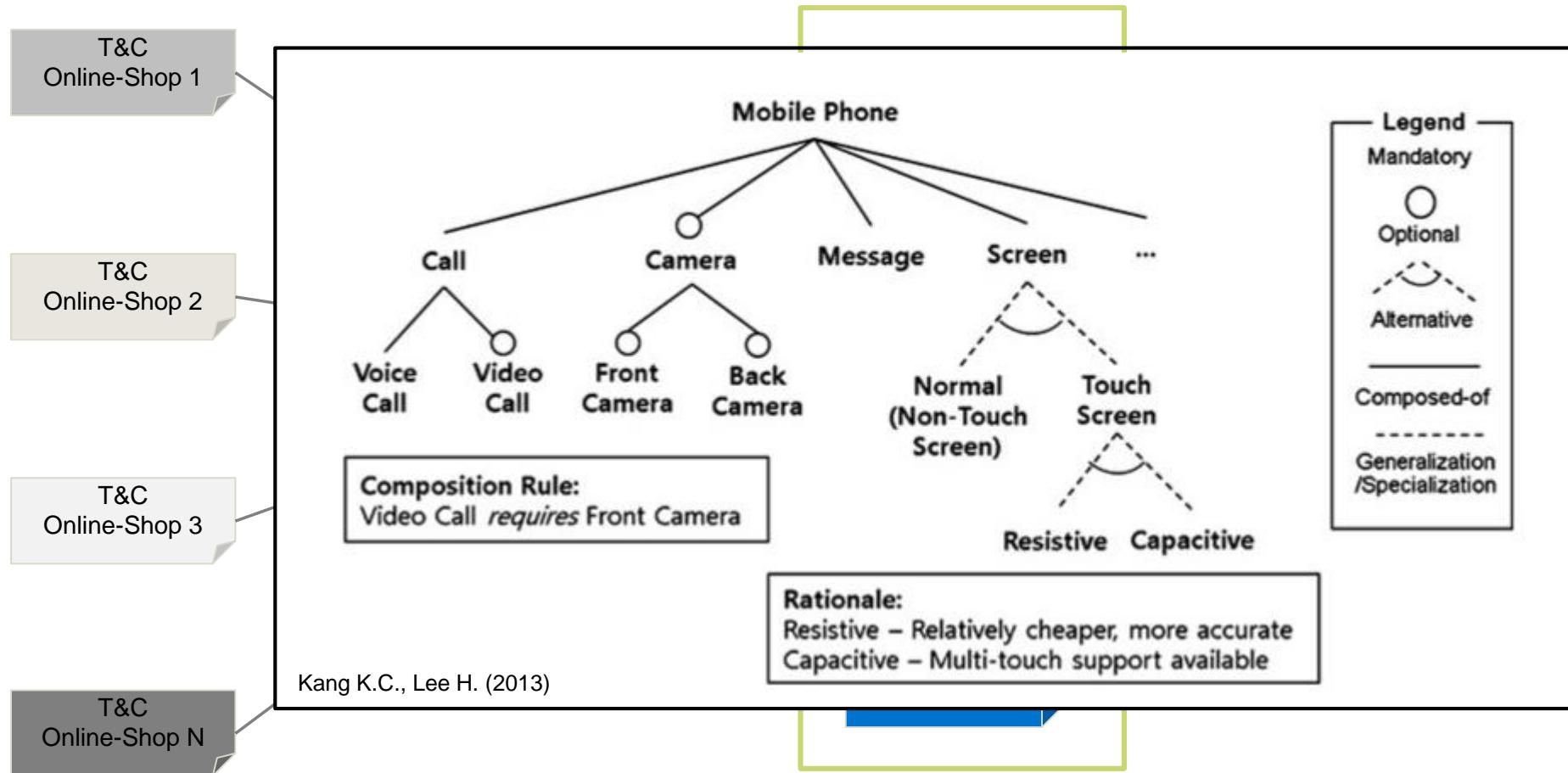


Goal: Creating a Variability Model of Terms & Conditions containing dependencies, necessary and unlawful clauses

Problem Statement



Problem Statement



Dataset

**Interactive
Machine Learning**

**Classified
Clauses**

**Interactive
Process**

**Variability
Model**

Introduction

- Motivation
- Problem Statement

Research Questions

- Interactive Machine Learning Introduction
- Data Collection and Preprocessing
- Text Classification

Approach

- Workflow
- Timetable

- How can **interactive Machine Learning** support clustering clauses from general Terms & Conditions?
- What are **suitable classes** and how can we classify clauses by their **semantic context**?
- How does a **prototypical implementation** enabling the semantic text matching of clauses in Terms & Conditions look like?
- How should the **User Interface** for the interaction between software and human expert look like?
- What is **Variability Modeling** and how can it help in observing relations between / lawfulness of clauses?

What is interactive Machine Learning?

“Machine learning with a human in the learning loop, observing the result of learning and providing input meant to improve the learning outcome.”⁽¹⁾

„Es handelt sich dabei um Algorithmen, die mit – teils menschlichen – Agenten interagieren und durch diese Interaktion ihr Lernverhalten optimieren können.“⁽²⁾

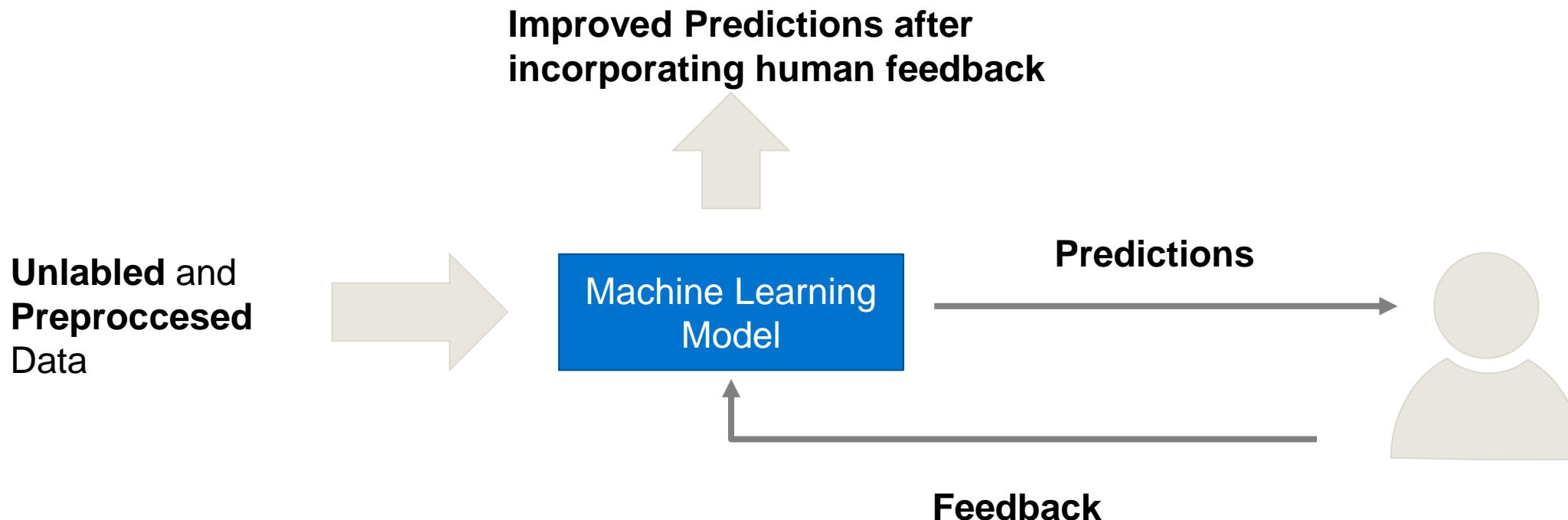
Why interactive ML?

- Interactive Machine Learning shows increase in both speed of learning and higher accuracy compared to other approaches.
- Solution with purely unsupervised approaches seems unfeasable

Analyzing current interactive Machine Learning (iML) solutions to identify possible approach

Example:

1. Starting with unsupervised learning
2. Human expert gives feedback on quality of predictions
3. Model incorporates feedback for new predictions



Building a usable Dataset: Data Collection

Crawling idealo.de for Online Shops

Based on a crawler from Daniel Braun



Extracting AGBs from Shops

Extracting the <body> of HTML form collected URLs and creating .txt File for each Online Shop (unfortunately containing irrelevant Information)



Manual Cleanup of collected Text

Manually removing irrelevant data, adding placeholders (e.g. for address of Online Shops), and general formatting

Important
No labeled Dataset available

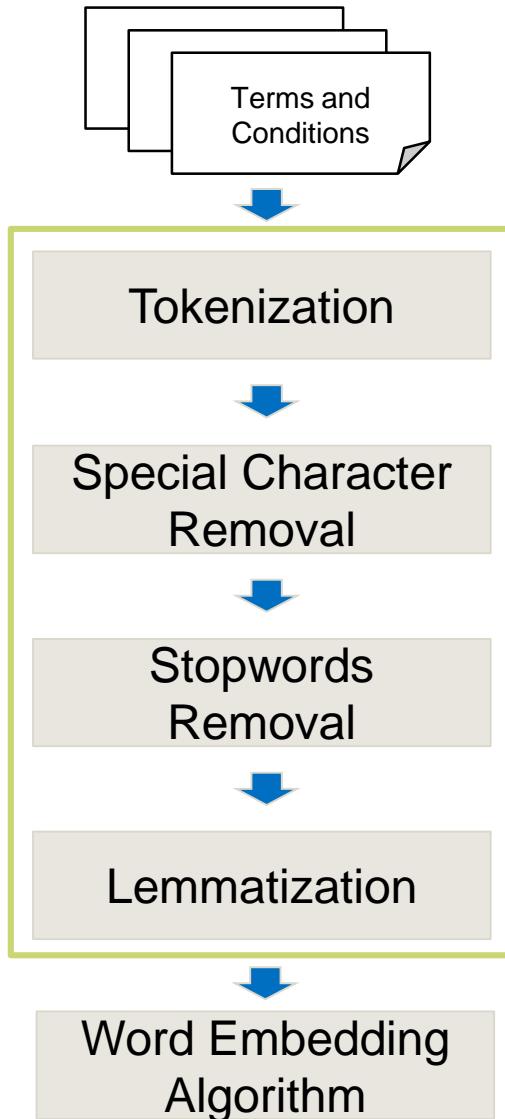
Creating .csv File containing general information and URLs for Terms and Conditions of partnered Online Shops

	A	B	C	D	E	F
1	id, name, shop_url, agb_url					
2	1, real GmbH, www.real.de, https://www.real.de/rechtliches/agb/					
3	2, comtech GmbH, www.comtech.de, https://www.comtech.de/unsere-agb					
4	3, Cyberport GmbH, www.cyberport.at, https://www.cyberport.at/cms/2/1/0/0/cm					
5	4, computeruniverse GmbH, www.computeruniverse.net, https://www.computeruni					
6	5, Handelshaus Wagner GmbH, www.elektroshopwagner.de, https://www.elektrosh					

Datenschutz
ElektroG
Akku- & Batterieentsorgung
Kundeninfos
Unternehmen Mindfactory
Allgemeine Geschäftsbedingungen
der Firma Mindfactory AG, Preußenstraße 14 a - c, 26388 Wilhelmshaven, nachstehend als Verkäufer bezeichnet
§ 1 Allgemeines, Begriffsbestimmungen
(1) Der Verkäufer bietet über den Online-Shop auf der Webseite www.mindfactory.de insbesondere Dienstleistungen an
(2) Verbraucher ist jede natürliche Person, die ein Rechtsgeschäft zu Zwecken abschließt,

1. Geltung
Diese allgemeinen Geschäftsbedingungen gelten für sämtliche Lieferungen der 21run GmbH (nachfolgend als Verkäufer bezeichnet).
2. Leistung
Über den Online-Shop können Sie Bekleidungs- sowie Sportartikel für den privaten Bedarf in handelsüblichen Mengen erwerben.
3. Vertragsprachen
Vertragsprachen sind Deutsch, Englisch, Spanisch und Französisch.
4. Vertragsabschluss / Bestellvorgang
Indem wir die Produkte in unserem Onlineshop einstellen, geben wir ein verbindliches Angebot zum Vertrag. Unmittelbar nach Absenden Ihrer Bestellung erhalten Sie per E-Mail eine Bestätigung über den Eingang Ihrer Vertragspartner ist die [[Anschrift]], eingetragen im Handelsregister beim Amtsgericht Mannheim. Ihre Bestelldaten können Sie nach Abgabe Ihrer Bestellung jederzeit über Ihr Kundenkonto einsehen.

Building a usable Dataset: Data Preprocessing



Die Lieferzeit beträgt, sofern nicht beim Angebot anders angegeben, 3-5 Tage.

['Die', 'Lieferzeit', 'beträgt', ',', 'sofern', 'nicht', 'beim', 'Angebot', 'anders', 'angegeben', ',', '3', '-', '5', 'Tage', '.']

['Die', 'Lieferzeit', 'beträgt', 'sofern', 'nicht', 'beim', 'Angebot', 'anders', 'angegeben', '3', '5', 'Tage']

['Die', 'Lieferzeit', 'beträgt', 'sofern', 'Angebot', 'anders', 'angegeben', '3', '5', 'Tage']

[,der', 'Lieferzeit', 'betragen', 'sofern', 'Angebot', 'anders', 'angeben', '3', '5', 'tagen']

Word2Vec, Doc2Vec etc.

Two possible levels of abstraction for classes:

Paragraph Level:

- Each class contains all clauses semantically belonging to one paragraph
- Much easier since paragraph title's contain valuable descriptors
- Few, but big Clusters (Total number of classes easier to predict: one for each type of Paragraph)
- Possible Classes: “§1 Grundlgende Bestimmung”, “§2 Zustandekommen des Vertrages” etc.

Clauses Level:

- Each class only contains clauses that are semantically the same
- Very hard, since wording in Terms & Conditions differs greatly, which makes finer classification difficult
- Many, but more precise Clusters (Total number hard to predict: “how many types of clauses exist?”, Overlap)
- Possible classes: “Vertragsgegenstand”, “Vertragsabschluss”

§ 1 Grundlegende Bestimmungen

1) Die nachstehenden Geschäftsbedingungen gelten für Verträge die über <http://www...> geschlossen werden. Soweit nicht anders vereinbart, wird eigens von Ihnen verwendeten Bedingungen widersprochen.

2) Verbraucher im Sinne dieser Regelung ist jede natürliche Person, die ein Rechtsgeschäft mit privaten Anliegen abschließt. Unternehmer ist jede natürliche oder juristische Person, die bei Abschluss des Rechtsgeschäfts im Sinne ihrer beruflichen oder gewerblichen Interessen handelt.

§ 2 Zustandekommen des Vertrages

1) Vertragsgegenstand ist der Verkauf von Waren.

2) Mit der Bereitstellung eines Produktes in unserem Shop unterbreiten wir unseren Kunden ein verbindliches Angebot zum Abschluss eines Kaufvertrages.

3) Alle zum Kauf beabsichtigten Produkte werden vom Kunden im „Warenkorb“ abgelegt. Nach Eingabe der persönlichen Daten und Zahlungsinformationen hat der Kunde die Möglichkeit alle eingegebenen Informationen zu überprüfen. Mit dem Absenden der Bestellung durch Klick auf die dafür vorgesehene Schaltfläche erklärt der Kunde rechtsverbindlich die Annahme des Angebotes. Damit ist der Kaufvertrag zustande gekommen. Bei der Zahlungs-Option über Payment-Dienstleister wie PayPal oder Sofortüberweisung, wird der Kunde von unserem Onlineshop auf die Webseite des Anbieters weitergeleitet. Nach Eingabe aller erforderlichen Daten wird der Kunde abschließend zurück in unseren Shop geleitet.

4) Die Übermittlung aller Informationen im Zusammenhang mit dem Vertragsschluss erfolgt automatisiert per E-Mail. Der Kunde hat daher sicherzustellen, dass die bei uns hinterlegte E-Mail-Adresse erreichbar ist.

Source: <https://website-tutor.com/agb-muster/>

Identification of suitable classes in context of Terms & Conditions

§ 1 Grundlegende Bestimmungen

- 1) Die nachstehenden Geschäftsbedingungen gelten für Verträge die über <http://www...> geschlossen werden. Soweit nicht anders vereinbart, wird eigens von Ihnen verwendeten Bedingungen widersprochen.
- 2) Verbraucher im Sinne dieser Regelung ist jede natürliche Person, die ein Rechtsgeschäft mit privaten Anliegen abschließt. Unternehmer ist jede natürliche oder juristische Person, die bei Abschluss des Rechtsgeschäfts im Sinne ihrer beruflichen oder gewerblichen Interessen handelt.

§ 2 Zustandekommen des Vertrages

- 1) Vertragsgegenstand ist der Verkauf von Waren.
- 2) Mit der Bereitstellung eines Produktes in unserem Shop unterbreiten wir unseren Kunden ein verbindliches Angebot zum Abschluss eines Kaufvertrages.
- 3) Alle zum Kauf beabsichtigten Produkte werden vom Kunden im „Warenkorb“ abgelegt. Nach Eingabe der persönlichen Daten und Zahlungsinformationen hat der Kunde die Möglichkeit alle eingegebenen Informationen zu überprüfen. Mit dem Absenden der Bestellung durch Klick auf die dafür vorgesehene Schaltfläche erklärt der Kunde rechtsverbindlich die Annahme des Angebotes. Damit ist der Kaufvertrag zustande gekommen. Bei der Zahlungs-Option über Payment-Dienstleister wie PayPal oder Sofortüberweisung, wird der Kunde von unserem Onlineshop auf die Webseite des Anbieters weitergeleitet. Nach Eingabe aller erforderlichen Daten wird der Kunde abschließend zurück in unseren Shop geleitet.
- 4) Die Übermittlung aller Informationen im Zusammenhang mit dem Vertragsschluss erfolgt automatisiert per E-Mail. Der Kunde hat daher sicherzustellen, dass die bei uns hinterlegte E-Mail-Adresse erreichbar ist.

Präambel

Die [Firma] (im Folgenden: [Firmenname ohne Gesellschaftsform]) stellt auf der Webseite [•] [Waren/Dienstleistungen], insbesondere [• und ähnliches] (im Folgenden: [allgemeine Artikelbezeichnung]) zur Verfügung.

Die Webseiten von [Firma] richten sich an [•] (im Folgenden: Nutzer).

Vertragsgegenstand

§ 1 Vertragsgegenstand

Gegenstand eines Vertrages eines Nutzers mit [Firma] sind [Dienstleistungen/Warenbestellungen] im Zusammenhang mit [Onlineshop] der [Firma], also [Artikelbezeichnung].

Vertragsabschluss

§ 2 Nutzung des Onlineshops, Vertragsschluss

Der Nutzer hat auf den Webseiten von [Firma] die Möglichkeit, [Artikelbezeichnung] und ähnliche Produkte gegen Entgelt zu erwerben.
Durch Anklicken der Schaltfläche „kaufen“ gibt der Nutzer ein rechtsverbindliches Angebot zum Kauf des angeforderten Produkts gegenüber [Firma] ab. [Firma] nimmt dieses Angebot durch Übersendung einer Email an.
Die Vertragsabwicklung erfolgt ausschließlich in deutscher Sprache.
Die Produktbeschreibung ist keine Garantie.

Paragraph
Clause

Sources: <https://website-tutor.com/agb-muster/>
<https://www.juraforum.de/muster-vorlagen/agb-online-shop>

Introduction

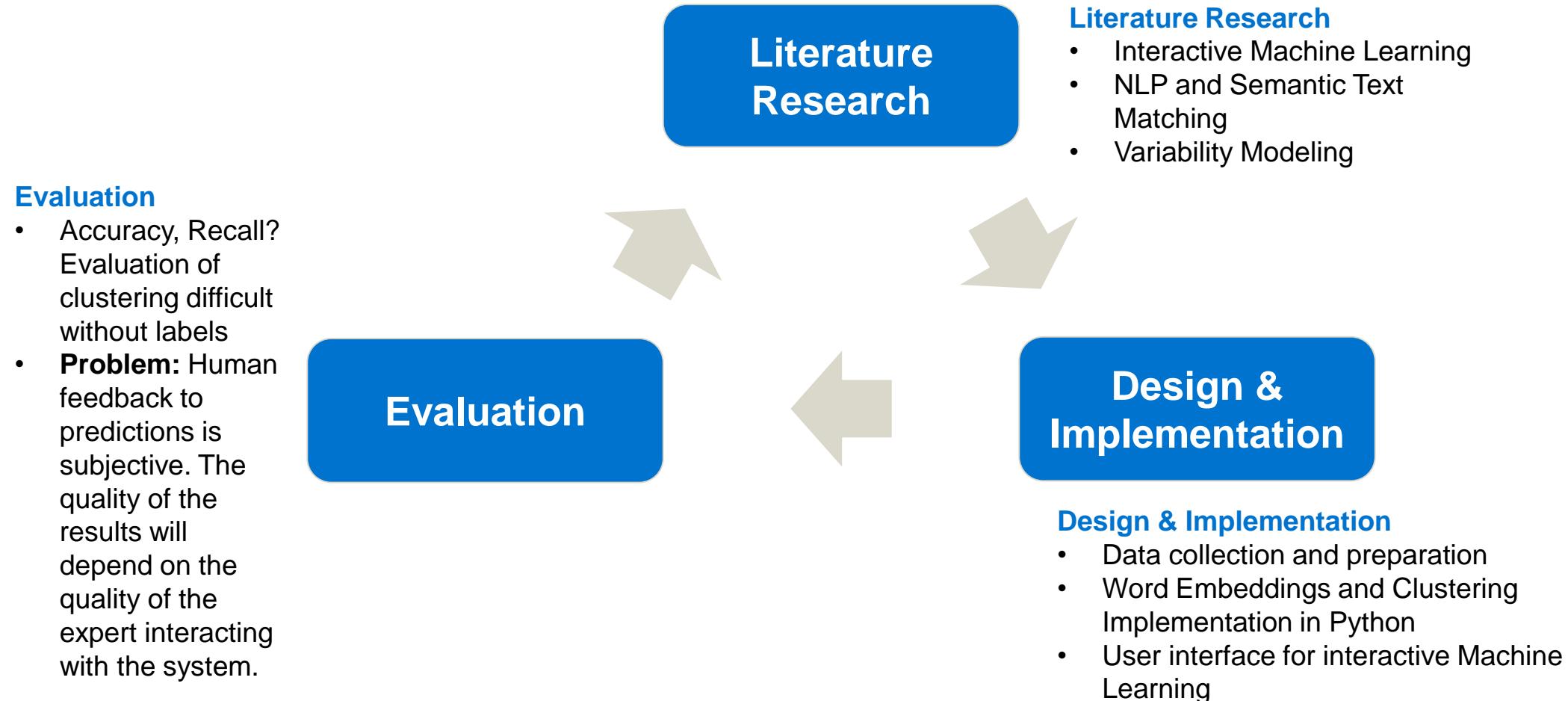
- Motivation
- Problem Statement

Research Questions

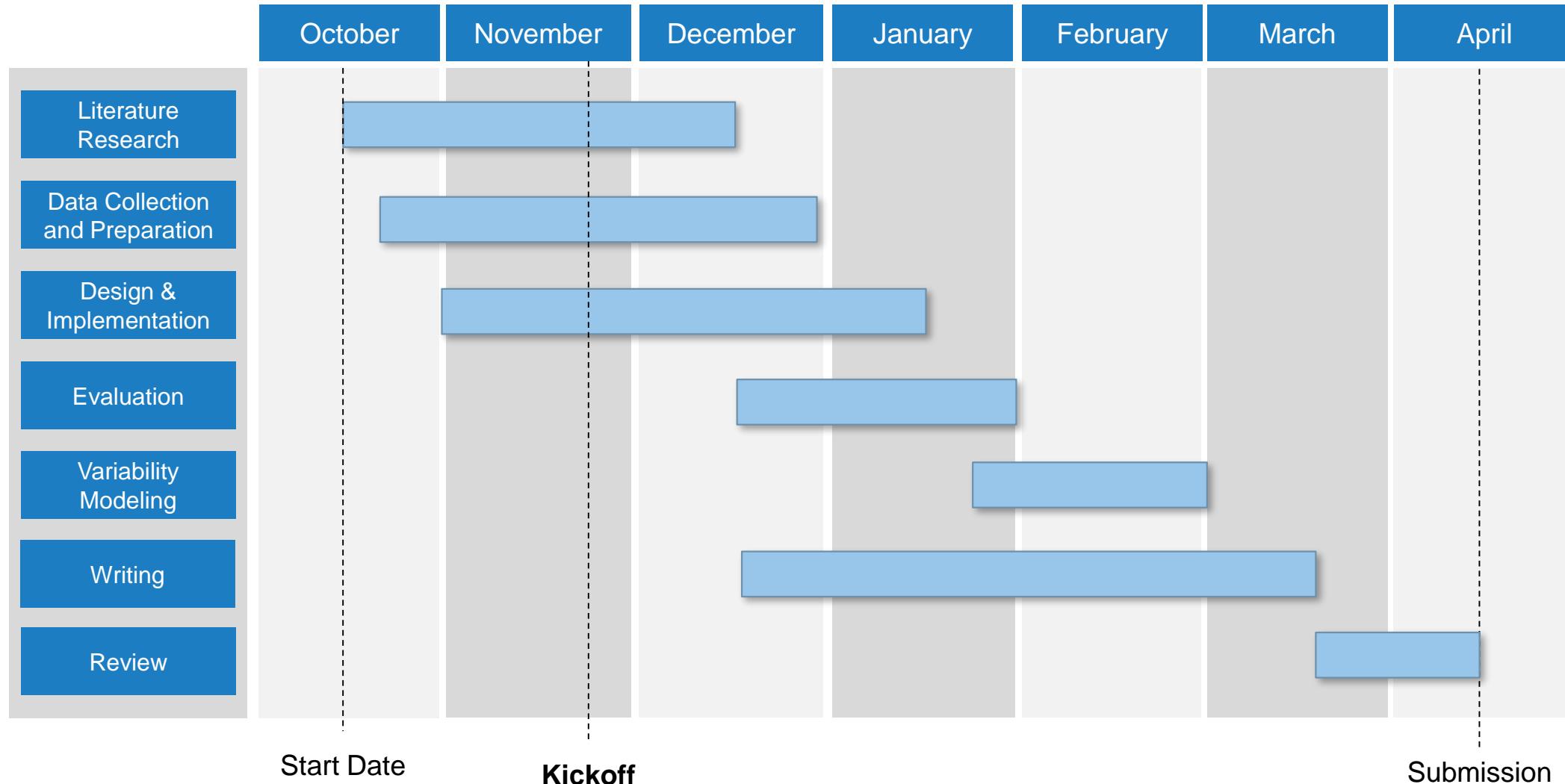
- Interactive Machine Learning Introduction
- Data Collection and Preprocessing
- Text Classification

Approach

- Workflow
- Timetable



Timetable





B. Sc.

David Koller

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289. 17132
Fax +49.89.289.17136

matthes@in.tum.de
wwwmatthes.in.tum.de



- Kang K.C., Lee H. (2013) Variability Modeling. In: Capilla R., Bosch J., Kang KC. (eds) Systems and Software Variability Management. Springer, Berlin, Heidelberg
- <http://iml.media.mit.edu/>
- Gesellschaft für Informatik: <https://gi.de/informatiklexikon/interactive-machine-learning-iml/>
- <https://website-tutor.com/agb-muster/>
- <https://www.juraforum.de/muster-vorlagen/agb-online-shop>