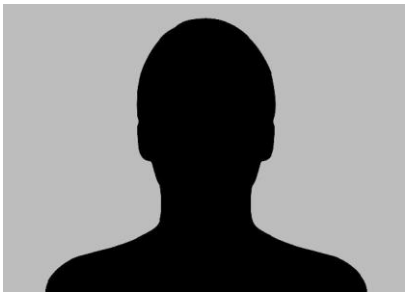
Zürcher Hochschule
für Angewandte Wissenschaften

School of
Engineering

InIT Institut für angewandte
Informationstechnologie

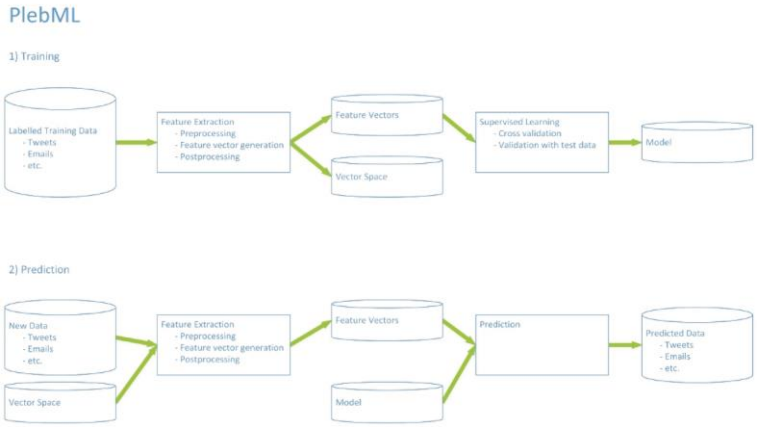
Ausbau eines Frameworks zur automatischen Textklassifikation



In der Software-Industrie werden Anwendungen des maschinellen Lernens seit längerer Zeit eingesetzt. Eines der Probleme dabei ist, dass dafür oft Spezialisten benötigt werden. Eine einfach zu verwendende und erweiterbare Bibliothek für maschinelles Lernen würde vielen Software-Entwicklern den Einstieg in diese Thematik erleichtern. In dieser Arbeit wurde eine Software-Bibliothek für maschinelles Lernen entwickelt, aufgeteilt in eine saubere Software-Architektur nach den Bedürfnissen des maschinellen Lernens. Zusätzlich wurden im Rahmen dieser Arbeit Interfaces mit Standard-Funktionen für einfache Textklassifikations-Aufgaben entwickelt und an mehreren Datensets getestet. Weiter wurden Anleitungen zur Erweiterung und Nutzung des Frameworks erarbeitet. Das Resultat ist ein robuster und erweiterbarer Kern, sowie ein domänenspezifisches Modul zur Textklassifikation.

Diplomierende
Marek Arnold
Dominic Egger

Dozierende
Mark Cieliebak
Fatih Uzdilli



Schematische Ansicht von den verschiedenen Schritten in PlebML.