

Alpha Analytics

KI im Handwerk

1. Einordnung

Für eine digitale Auftragsverarbeitung im Handwerk müssen zu Beginn alle relevanten Informationen eines Auftrags im Softwaresystem erfasst werden. Eine Teilaufgabe in dem Bereich ist das Anlegen der Tätigkeiten, die im Rahmen eines Auftrags anfallen. Diese müssen durch manuelles Lesen und Zuordnen angelegt werden. Dieser Schritt ist zeitaufwändig und sinnvoll zu automatisieren. Zielstellung des Projektes war es, diese Automatisierung durch Auswertung historischer Auftragsdaten und durch Einsatz maschinellen Lernens zu unterstützen.

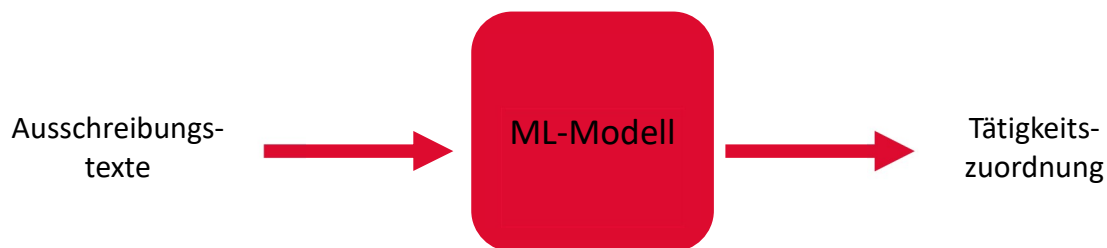


Abbildung 1: Einsatz eines ML-Modells

Dabei sollten einzelne Beschreibungstexte aus Ausschreibungen eingelesen, von einem trainierten Machine Learning Modell verarbeitet und eine oder mehrere Tätigkeiten daraus abgeleitet werden. Diese stehen dann als Liste für die Nutzung im Rahmen der digitalen Auftragsverarbeitung zur Verfügung. Abbildung 1 zeigt die grundlegende Funktionsweise schematisch.

2. Datengrundlage: Datenschutz im Handwerk

Einer der wichtigsten Punkte für den sinnvollen Einsatz maschinellen Lernens ist das Vorhandensein ausreichender Daten und Datenmengen. Dies beinhaltet zum einen wortwörtlich ausreichend viele Daten, zum anderen aber auch ausreichend Daten im Hinblick auf die zu lösende Fragestellung. In diesem Fall heißt das: ausreichend Beispiele von Ausschreibungstexten und der passend zugeordneten Tätigkeiten. Für diese Datenlage können dann entsprechend überwachte Lernverfahren für das Training robuster und leistungsfähiger ML-Modelle eingesetzt werden.

Im vorliegenden Projekt wurden Daten aus mehr als 10 Jahren digitaler Auftragsdatenverarbeitung zur Verfügung gestellt und verarbeitet. Dabei wurden über 32.000 Texte und ihre Zuordnung zu einer Tätigkeitsklasse verarbeitet, welche aus über 400 Ausschreibungen stammen. Texte mit mehrfacher Klassenzugehörigkeit sind entsprechend redundant im Gesamtdatensatz angelegt. Darüber hinaus wurden weitere 500 Ausschreibungstexte in die Datenvorverarbeitung integriert. Ein solcher Datensatz in diesem Bereich repräsentiert viele Jahre Erfahrung und stellt eine sehr gute

Alpha Analytics

Voraussetzung dar, um diese Erfahrung mit maschinellem Lernen zu erfassen, zu kapseln und nutzbar zu machen im Rahmen eines automatisierten Einsatzes.

Anteil Tätigkeiten an Objekt ID 944

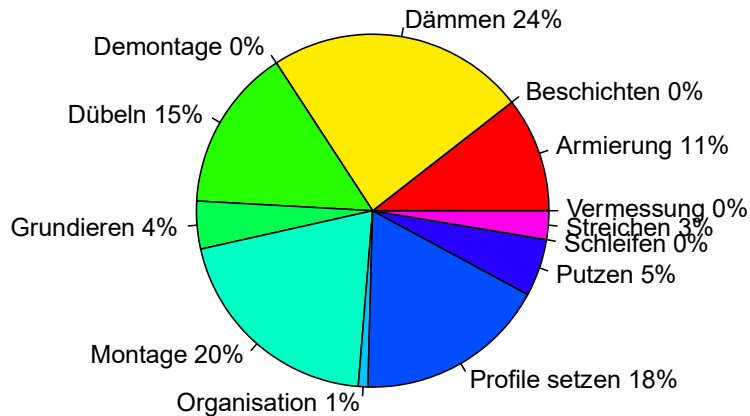


Abbildung 2: Anteil der untersuchten Tätigkeiten beispielhaft für ein Objekt

Abbildung 2 zeigt beispielhaft für einen Auftrag, der aus zahlreichen Einzeltexten besteht, zu welchen Anteilen typischen Tätigkeiten im Rahmen eines Auftrags vorkommen. Die Verteilung in einem beispielhaften Auftrag für ein bestimmtes Objekt lässt bereits vermuten, dass die Tätigkeiten auch nicht gleichmäßig verteilt im Gesamtdatensatz vorkommen: Ein wichtiger Punkt, der das Modelltraining und die spätere Validierung des Modells bzw. dessen Leistungsfähigkeit beeinflusst.

3. Datenaufbereitung

Ein wichtiger Punkt im Rahmen der Entwicklung eines passenden NLP-Modells ist die Aufbereitung der Daten, um Charakteristika für das Training zugänglich zu machen und irrelevante Datenbereiche zu entfernen. Dabei wurden u.A. Standardansätze wie die Reduktion der Wörter auf ihre Wortstämme genutzt. Im Rahmen der Aufbereitung wurden auch verschiedene statistische Eigenschaften der untersuchten Texte berücksichtigt. Des Weiteren wurde mit Part-Of-Speech-tagging (POS-Tagging) Substantive und Verben erkannt, um relevante für die weitere Verarbeitung und Klassifizierung identifizieren zu können. Durch das POS-Tagging konnten weitere Informationen über die Zusammensetzung des Textmaterials gewonnen werden. Dabei wurden im Schnitt 3,6 Sätze bzw. 59 Wörter für einen durchschnittlichen Einzeltext in einer Ausschreibung bestimmt. Abbildung 3 visualisiert die Details.

Alpha Analytics

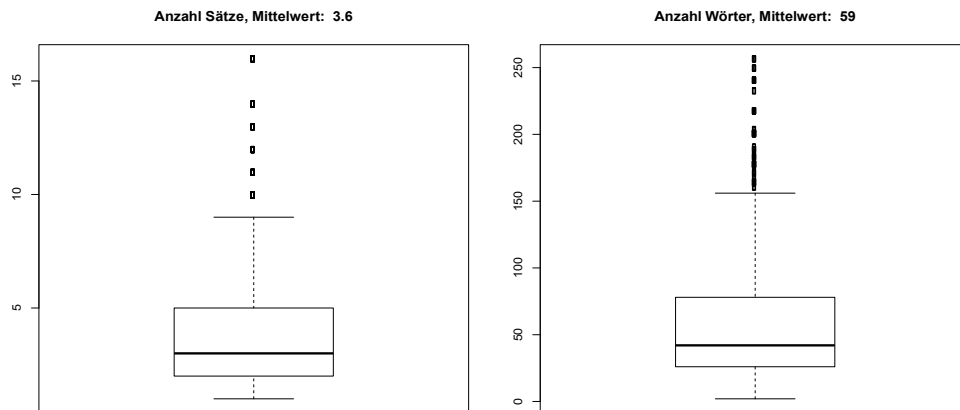


Abbildung 3: Durchschnittliche Anzahl an Wörtern bzw. Sätzen in einem Einzeltext

Daneben zeigt Abbildung 4 die durchschnittliche Anzahl von Substantiven und Verben über verschiedene Aufträge hinweg. Allgemein waren etwas abweichende Charakteristika in den Ausschreibungstexten im Vergleich zu Standardtexten aus Zeitungen und Medien wie Wikipedia erkennbar. Hier könnte ein zusätzlicher Grund für das vergleichsweise schlechtere Abschneiden der verfügbaren Standardmodelle vermutet werden (neben den zusätzlichen Anforderungen solcher Modellansätze an verfügbare Datenmengen).

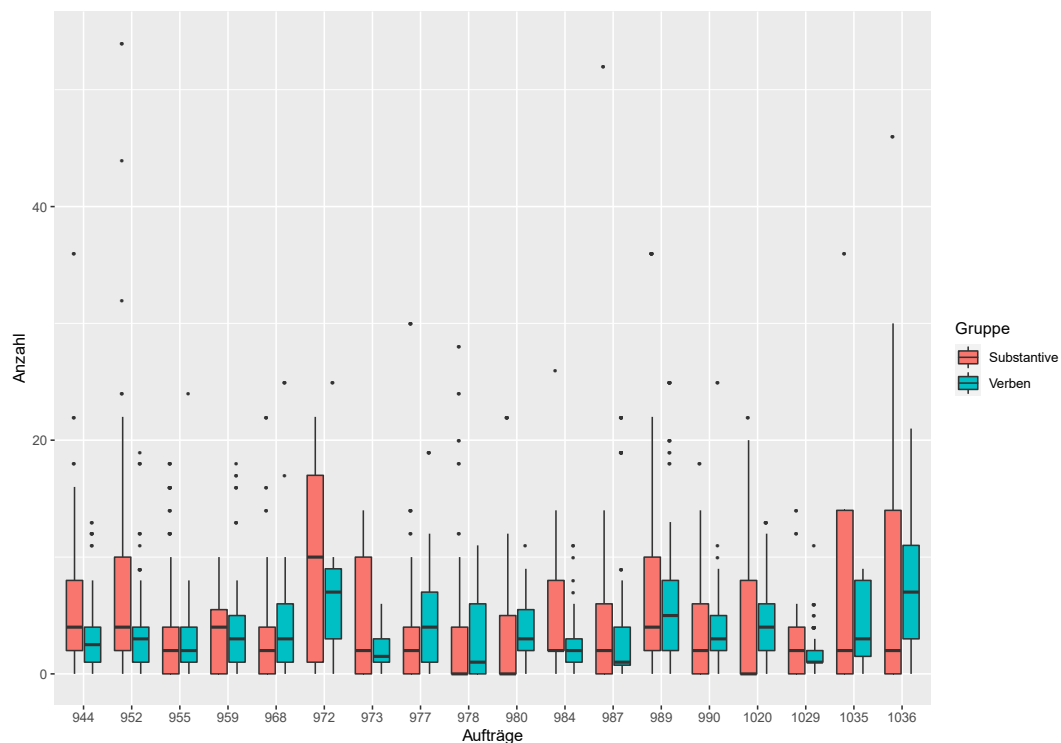


Abbildung 4: Durchschnittliche Anzahl Substantive bzw. Verben in verschiedenen Aufträgen

Alpha Analytics

Entsprechend wurden für die weitere Verarbeitung die sogenannte word embedding oder Worteinbettung genutzt. Dabei werden Wörter mit Hilfe von Vektoren dargestellt und ermöglichen die Abbildung verschiedener Kontextinformationen. Diese Vorverarbeitung war eine wichtige Grundlage für das nachfolgende Modelltraining.

4. Modellentwicklung und Funktionsweise

Für die Umsetzung der Aufgabenstellung, Tätigkeiten aus Ausschreibungstexten ableiten zu können, wurde ein Klassifikationsmodell angestrebt. Dabei stellen die verschiedenen Tätigkeiten die zu erkennenden Klassen dar.

Im ersten Ansatz wurden vortrainierte, frei verfügbare NLP-Modelle (openNLP) genutzt. Diese konnten jedoch einzelne Klassen nicht ausreichend gut auseinanderhalten. Daher konnten entsprechende Modelle nicht zufriedenstellend eingesetzt werden. Entsprechend wurden eigen entwickelte Modellansätze verfolgt. Erste Varianten waren auch im Bereich RNN (Recurrent Neural Network) bzw. LSTM (Long Short-Term Memory) angesiedelt. Die verfolgten Ansätze konnten nicht robust trainiert werden und waren schlussendlich nicht ausreichend leistungsfähig. Damit konnten klassische Ansätze im Bereich NLP nicht genutzt werden, um die zugrundeliegende Aufgabe der Tätigkeitklassifikation zufriedenstellend zu lösen.

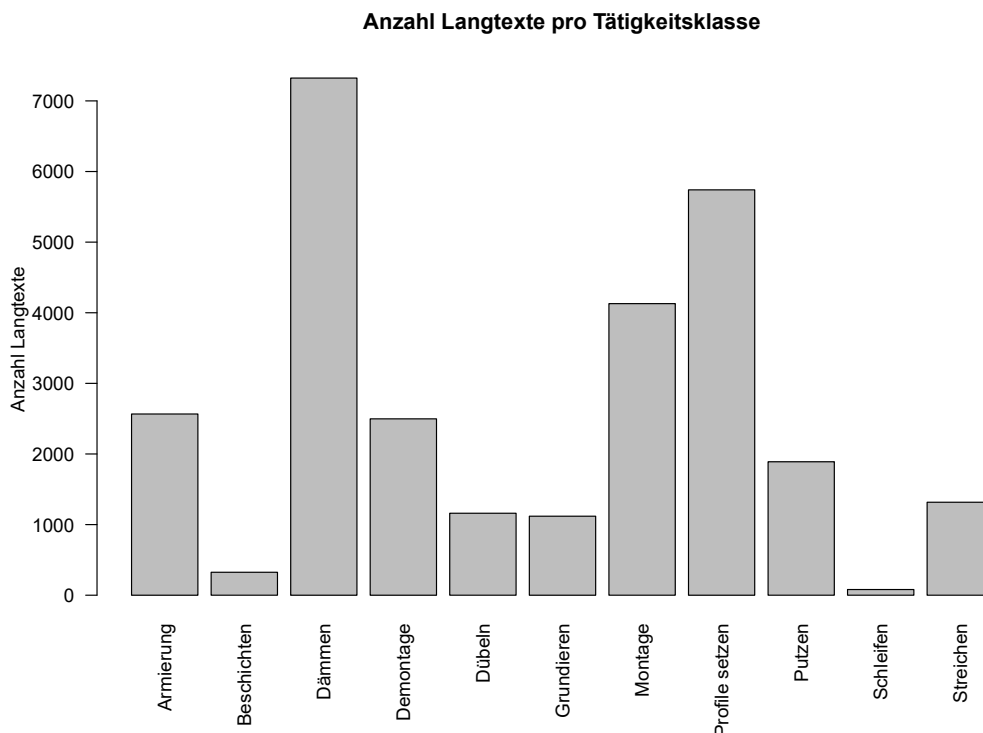


Abbildung 5: Verteilung der Datenmenge auf die Tätigkeitsklassen

Alpha Analytics

Nach verschiedenen Aufbereitungsschritten und der Berücksichtigung der Text-Charakteristika wurden Ensemble Learning Methoden herangezogen und basierend auf den Daten ein Random Forest trainiert. Beim Training solcher Modelle ist die Balance der Daten über die verschiedenen Klassen sehr wichtig. Im Idealfall werden die Klassen durch gleich große Datenmengen/Einträge repräsentiert. Abbildung 5 zeigt die Verteilung der Daten über die realen Tätigkeits-Klassen in diesem Projekt. Dabei wird deutlich, dass die zugrundeliegenden Daten nicht gleichmäßig über alle Klassen verteilt sind, sondern zwischen verschiedenen Tätigkeitsklassen schwanken. Die Klasse mit dem größten Datenumfang ist die Tätigkeit „Dämmen“, die Klasse mit dem kleinsten Umfang ist „Schleifen“, dicht gefolgt von der Tätigkeitsklasse „Beschichten“. Dies muss bei einzelnen Modelltypen wie z.B. Random Forests in Training und Validierung entsprechend berücksichtigt werden.

Für die 11 Tätigkeitsklassen konnte ein Random Forest trainiert werden, der eine durchschnittliche Genauigkeit von 64% und eine Klassen-Sensitivität von 68% erreicht hat. Abbildung 6 zeigt die erreichte Sensitivität für jede einzelne Tätigkeitsklasse. Trotz der Berücksichtigung der Unterschiede im Datenumfang der einzelnen Tätigkeitsklassen beim Modelltraining bestanden starke Unterschiede in der Klassifizierungsfähigkeit des Modells in Bezug auf die einzelnen Klassen.

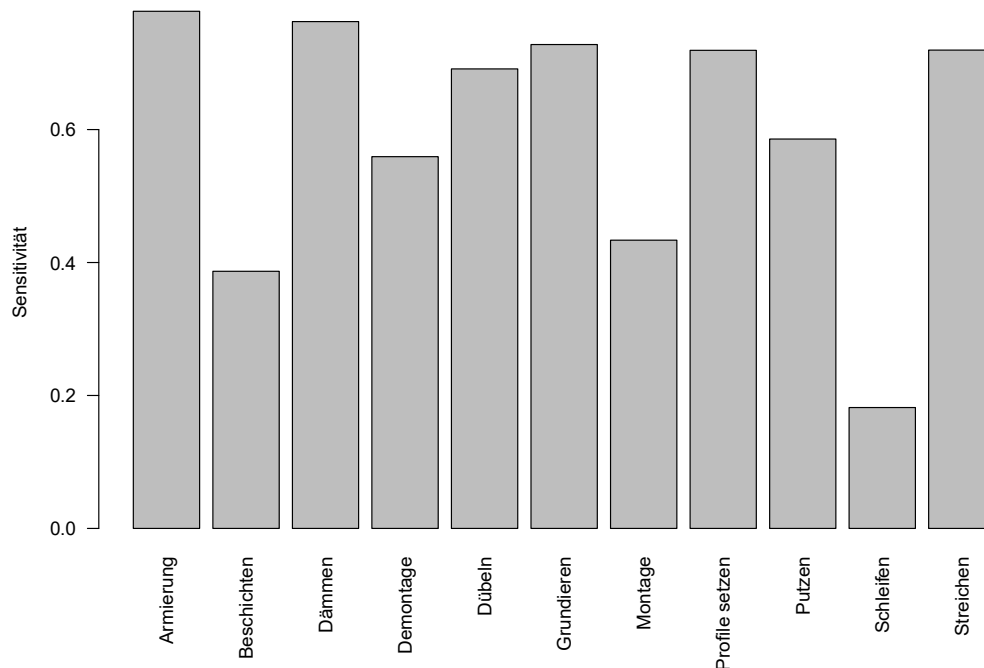


Abbildung 6: Sensitivität für Einzelklassen

Mit Blick auf eine Stabilisierung des Modells wurden die beiden kleinsten Klassen nach Untersuchung ihrer tatsächlichen Auftrittshäufigkeit über verschiedene Aufträge und in Absprache mit den Domänenexperten aus dem Klassifikationsansatz entfernt. Damit verbleiben neun Tätigkeitsklassen für die weitere Modellentwicklung. Auf dieser Basis konnte eine Abschätzungsgenauigkeit von ca. 72% erreicht und die Schwankungsbreite merkbar verkleinert werden.

Alpha Analytics

5. Single vs. Multiclass

Eine besondere Herausforderung im Rahmen der Tätigkeitsklassifikation ist die Erkennung von einer vorher nicht fest definierten Anzahl an Klassen / Tätigkeiten (1 bis typischerweise 4) für einen durchschnittlichen Textabschnitt / Aufgabenabschnitt.

Ein Beispiel eines solchen Textabschnitts, aus dem mehrere Tätigkeiten abzuleiten sind, ist im Folgenden gegeben (aus ObjektID 955):

Verputzte Außenwärmedämmung an Decke, Untergrund Beton, Dämmstoff aus Mineralwolle MW DIN EN 13162, Bemessungswert der Wärmeleitfähigkeit max. 0,035 W/(mK), einlagig, Dämmschichtdicke 120 mm, als Platte, stumpf gestoßen, Anwendungsgebiet DIN 4108-10 WAP, hohe Zugfestigkeit - zh, geklebt und konstruktiv gedübelt mit Schlagdübeln, 4 Stück/m², Dübellänge entsprechend Dämmstoffdicke und Dübelanzahl entsprechend Verankerungsgrund und Standsicherheitsnachweis, Armierungsputz aus mineralischem Werk trockenmörtel, Armierungsputz Dicke 3 bis 5 mm, einschl. Armierungsgewebe, Oberputz Kalkzementputz P II, einschl. Grundierung, als geriebener Putz, Körnung 3 mm. Hinweis zur Ausführung Decke im Eingangsbereich EG, im überdeckten Außenbereich ohne direkte Bewitterung, Ausführung Armierungsputz, -gewebe und Oberputz wie an Wandflächen

Dieser Text steht für vier auszuführende Tätigkeiten und damit vier Klassen, die im Text erkannt werden müssen: **Grundieren (218) / Dämmen (258/259) / Dübeln (289) / Putzen (316)**. Eine solche Multiklassenerkennung ist nicht trivial möglich mit Standard-Modellen. Um den Anforderungen einer Mehrfacherkennung gerecht zu werden, wurde das einfache Modell für 11 bzw. 9 Klassen zu einem Modellsystem aus 9 Modellen ersetzt, die zusammenwirken, um eine entsprechende Mehrklassenabschätzung zu ermöglichen. Im neuen Modellsystem ist ein Modell im System für eine einzelne Klasse „zuständig“. Um eine Mehrklassenentscheidung zu ermöglichen, werden die Ergebnisse entsprechend konsolidiert.

Im Ergebnis umfasst die Abschätzung des Modellsystems eine Liste der Tätigkeiten und eine entsprechende Wahrscheinlichkeit für eine Klassenzugehörigkeit. Damit ist auch die notwendige Möglichkeit gegeben, eine flexible Anzahl an Klassen zuverlässig zu erkennen.

Für dieses Modellsystem konnte eine durchschnittliche Genauigkeit von 78% erreicht werden. Entsprechend zeigt Abbildung 7 zum Vergleich die Sensitivität, die mit dem Modellsystem durchschnittlich pro Tätigkeitklasse erreicht werden konnte. Diese liegt im Gesamtdurchschnitt bei ca. 72%. Über die 9 Klassen hinweg konnten deutlich gleichmäßigere Ergebnisse bei der Abschätzung und ein robusteres Modell erzielt werden.

Alpha Analytics

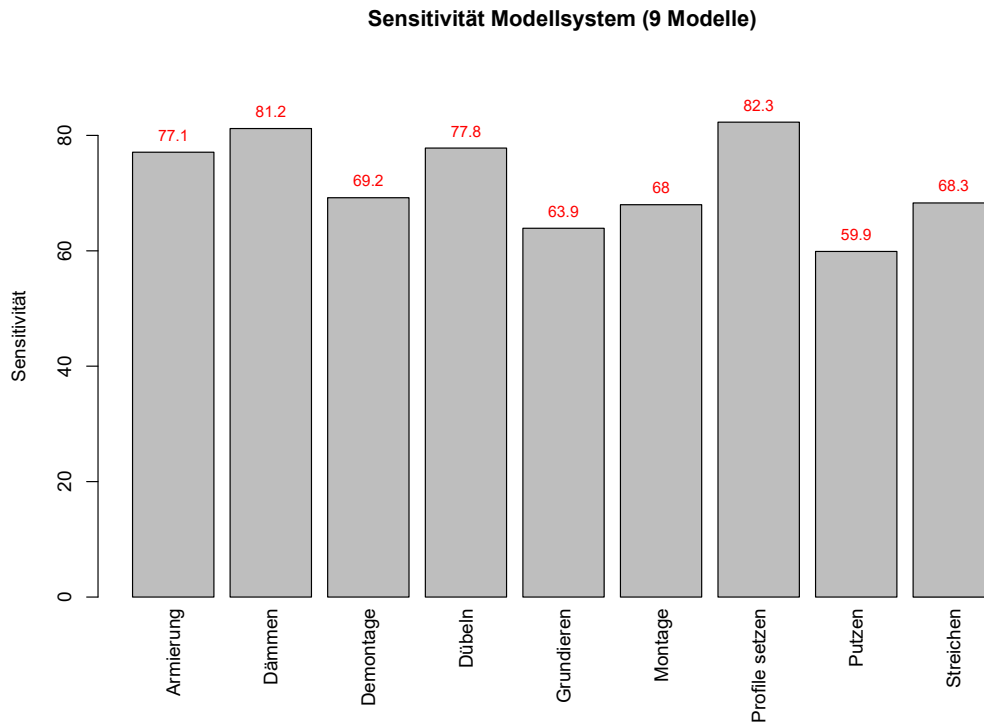


Abbildung 7: Sensitivität Klassenerkennung im Rahmen der Mehrklassenerkennung

6. Verlässlichkeitsprüfung

Nach dem Training gilt es die zu erwartende Leistungsfähigkeit des trainierten Modellsystems im späteren Einsatz abzuschätzen, um den sinnvollen Einsatz eines Modells realisieren zu können. Dafür können verschiedene Methoden und Metriken zum Einsatz kommen, die passende Wahl richtet sich im Idealfall nach dem zugrundeliegenden Anwendungsfall. Dies entspricht hier dem Level eines neuen Textes aus einem neuen Auftrag / aus einer neuen Ausschreibung, mit dem das Modellsystem validiert werden kann.

Im Projekt wurde mit einem unabhängigen Datensatz (Weissenstadt) eine entsprechende Validierung vorgenommen. Abbildung 8 und Abbildung 9 zeigen die Ergebnisse dieser unabhängigen Validierung sowie die Schwankungen der Genauigkeit (Balanced Accuracy) für jede Tätigkeit. Die erreichte Güte des Modellsystems ist ausreichend hoch für den praktikablen Einsatz im Rahmen der digitalen Auftragsverarbeitung.

Alpha Analytics

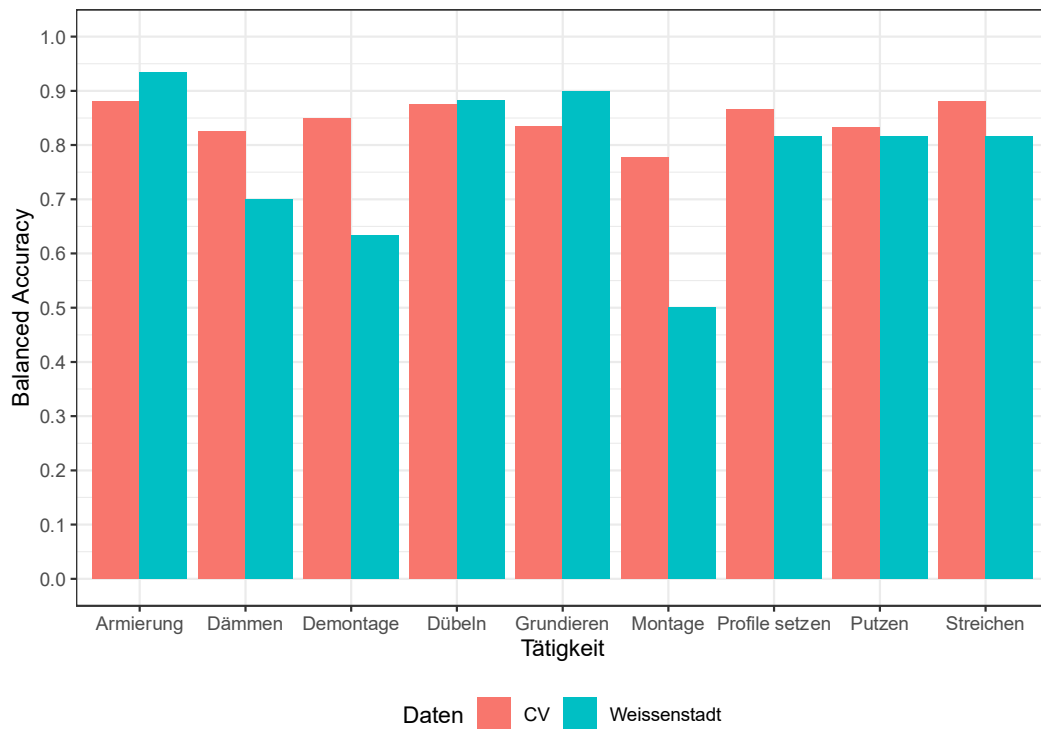


Abbildung 8: Validierungsergebnisse mit Test-Daten "Weissenstadt"

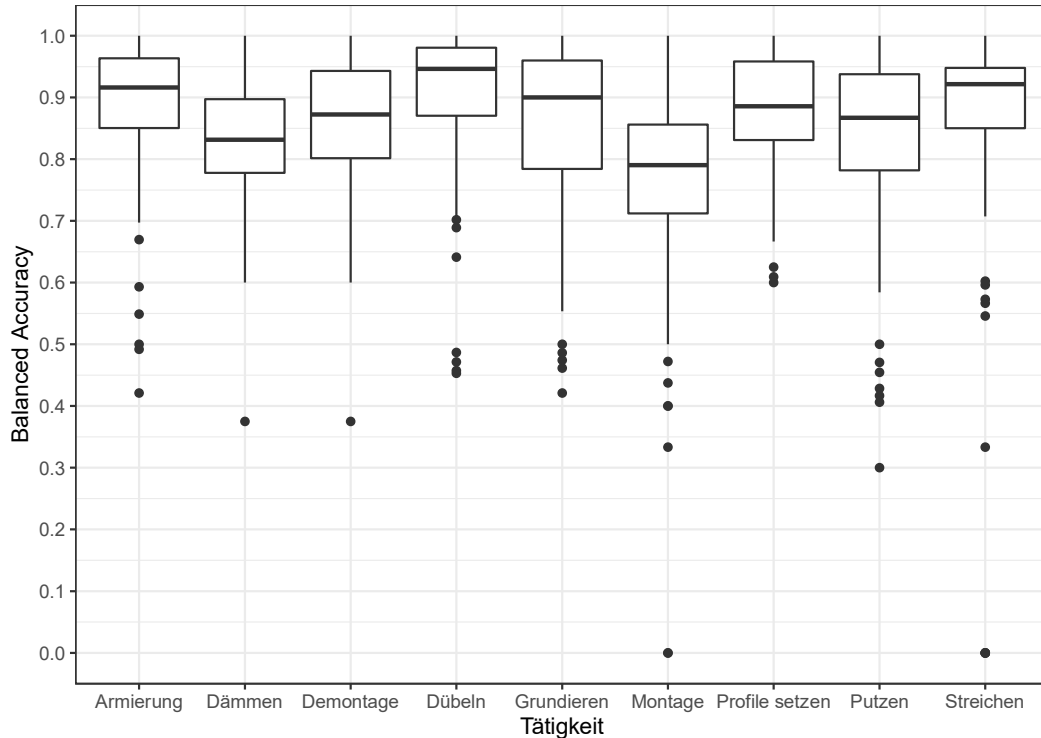


Abbildung 9: Übersicht Ergebnisse Kreuzvalidierung

Alpha Analytics

Nach Inbetriebnahme und Integration in den Softwareprozess ist eine regelmäßige, automatisierte Überwachung der initialen Leistungsfähigkeit von großem Vorteil, um eine gleichbleibende Leistungsfähigkeit und damit Prozessqualität des automatisierten Prozessschrittes zu gewährleisten.

7. Übertragbarkeit

Gerade im Hinblick auf den Einsatz des entwickelten Modells nicht nur in einem Handwerks-Fachbereich, sondern für verschiedene Gewerke ist wichtig, dass das trainierte Modell aktuell eine entsprechende Spezialisierung hat, da es auf Daten eines Gewerks trainiert wurde. Daher kann es nicht trivial und ohne Anpassungen mit der gleichen Leistungsfähigkeit für andere Gewerke eingesetzt werden.

Soll eine Übertragung ermöglicht werden, muss ein entsprechendes Nachtraining des Modells mit den Gewerke spezifischen Daten stattfinden. Dies kann auch nach und nach geschehen, während das Modell für Gewerke unspezifische Aufgaben teilweise zum Einsatz kommen kann. Im Rahmen eines nach und nach erfolgenden Nachtrainings kann mit einer angebundnen automatisierten Validierung und Leistungsbewertung gut abgeschätzt werden, wann ein nachtrainiertes Modell ausreichend gut funktioniert, um sich auf dessen Abschätzungen im neuen Wirkbereich verlassen zu können.

8. Implementierung und Einbettung

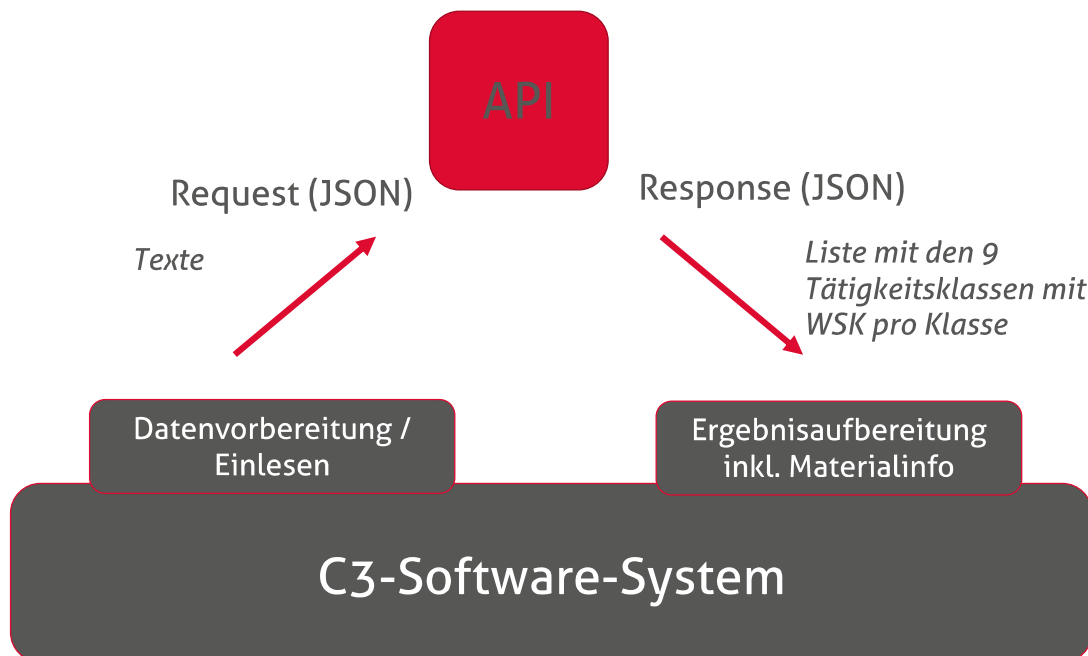


Abbildung 10: Einbettung in bestehende Softwaresysteme

Alpha Analytics

Die Implementierung des entwickelten Modellsystems wurde aus mehreren Random Forest Einsätzen in Form einer per Endpoint ansprechbaren API in Go umgesetzt. Damit kann das trainierte und validierte Modellsystem gekapselt und für den lokalen Einsatz aufbereitet werden. Abbildung 10 stellt schematisch die Einbettung der implementierten API in das C3-Softwaresystem dar. Das trainierte Modellsystem ist so eingebunden, dass es auf einfache Art und Weise mit einem nachtrainierten Modellsystem aktualisiert werden kann, ohne die vollständige API auszutauschen.

Die API wurde erweitert durch einen Validierungs-Endpoint für die regelmäßige Prüfung der aktuellen Leistungsfähigkeit des eingesetzten Modellsystems.

Das implementierte Modell (v.2.0) und die entsprechende Funktionalität stehen unter <https://nlp.alpha-analytics.de/> als Demonstrator zur Verfügung. Abbildung 11 zeigt den funktionalen Demonstrator mit einem Beispieltext und die dazugehörige Modellbewertung. Dieser kann auf der angebenen Website auch mit beliebigen, eigenen Texten getestet werde.

KI-Modell-Demonstrator
Hier kann das entwickelte KI-Modell als Demonstrator getestet werden.

Start
Ziehen Sie eine Kachel mit Beispieltext (T1/T2) das Feld rechts daneben.
Oder tragen Sie dort Ihren eigenen Ausschreibungstext ein, den Sie von der KI bewerten lassen wollen

KI bei der Arbeit
Sehen Sie in der Übersicht rechts, welche Tätigkeiten die KI im Ausschreibungstext verrichtet.
Umso länger der Balken im Diagramm, umso sicherer ist sich die KI für die angegebenen Tätigkeiten

Prüfung Leistungsfähigkeit der KI
Grün: Abschätzung des Modells stimmt mit dem Expertenwissen überein
Rot: Abschätzung stimmt nicht überein, wie z.B. für Text T3 aus einem anderen Bereich

Input Text:
T1: Grundierung stark saugender Untergründe mit emissionsarmen lösemittel- und weichmacherfreien, wasserverdünnbaren Tiefgrund. Emissionsarm, lösemittelfrei, geruchsneutral. Untergrund: Stahlbeton / Mauerwerk Fabrikat der Planung: Stoß Tiefgrund oder gleichwertig
T2: [Empty]
T3: [Empty]

Activity Classification Bar Chart:

Tätigkeit	Wahrscheinlichkeit Zuordnung
Montage	~0.10
Demontage	~0.10
Streichen	~0.10
Putzen	~0.10
Armierung	~0.10
Dübeln	~0.25
Dämmen	~0.10
Profile setzen	~0.10
Grundieren	1.00

Performance Indicators:
Three circles representing model performance: one green (T1), one white (T2), and one white (T3).

Abbildung 11: Demonstrator