

Machine Learning In Microsoft Azure - Ein Vergleich

DIPLOMARBEIT

verfasst im Rahmen der

Reife- und Diplomprüfung

an der

Höheren Abteilung für Informatik

Eingereicht von:

Polleichtner Moritz
Sljivic Emina

Betreuer:

Karpowicz Michał

Projektpartner:

Bammer Patrick, smartpoint IT consulting GmbH

Leonding, April 2022

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Die Arbeit wurde bisher in gleicher oder ähnlicher Weise keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Die vorliegende Diplomarbeit ist mit dem elektronisch übermittelten Textdokument identisch.

Leonding, April 2022

M. Polleichtner & S. Emina

Zur Verbesserung der Lesbarkeit wurde in diesem Dokument auf eine geschlechtsneutrale Ausdrucksweise verzichtet. Alle verwendeten Formulierungen richten sich jedoch an alle Geschlechter.

Abstract

Brief summary of our amazing work. In English. This is the only time we have to include a picture within the text. The picture should somehow represent your thesis. This is untypical for scientific work but required by the powers that are. Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Zusammenfassung

Zusammenfassung unserer genialen Arbeit. Auf Deutsch. Das ist das einzige Mal, dass eine Grafik in den Textfluss eingebunden wird. Die gewählte Grafik soll irgendwie eure Arbeit repräsentieren. Das ist ungewöhnlich für eine wissenschaftliche Arbeit aber eine Anforderung der Obrigkeit. *Bitte auf keinen Fall mit der Zusammenfassung verwechseln, die den Abschluss der Arbeit bildet!* Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Inhaltsverzeichnis

1	Machine Learning	1
1.1	Theorie	1
1.1.1	KI, ML, DL - Unterschiede	1
1.1.2	Arten von Machine Learning	3
1.1.3	Optical Character Recognition	9
2	Cloud Computing	14
2.1	Theorie	14
2.1.1	Was ist Cloud Computing?	14
2.1.2	Vorteile von Cloud Computing	15
2.1.3	Hindernisse für den Einsatz einer Cloud	15
2.1.4	Cloud Computing Modelle	16
2.1.5	Anwendungsfälle von Cloud Computing	17
2.2	Praxis: AI Builder	19
2.2.1	Benutzerdefinierte Modelle	19
2.2.2	AI-BUILDER in der Praxis	22
2.3	Praxis: Power Automate	24
2.3.1	Cloud Flow	24
2.4	Praxis: Cognitive Services	28
2.4.1	Cognitive Computing	30
2.4.2	Cognitive Computing und Cloud Computing	30
2.4.3	Microsoft Azure Formularerkennung	31
2.4.4	Formularerkennung	33
3	Ausgangslage	41
3.1	Ausgangssituation	41
3.2	Istzustand	41
3.3	Problemstellung	41

3.4	Ziele	41
3.5	Aufgabenstellung	41
3.5.1	Funktionale Anforderungen	41
3.5.2	Nicht funktionale Anforderungen	41
3.6	Systemarchitektur	41
3.7	Ablauf	41
4	Umsetzung	42
4.1	Konfigurations- und Entwicklungszeit	42
4.2	Performance und Präzision	42
4.3	Benutzerfreundlichkeit	42
4.4	Kosten	42
5	Gegenüberstellung und Conclusio	43
5.1	Konfigurations- und Entwicklungszeit	43
5.2	Performance und Präzision	43
5.3	Benutzerfreundlichkeit	43
5.4	Kosten	43
	Literaturverzeichnis	VI
	Abbildungsverzeichnis	VII
	Tabellenverzeichnis	VIII
	Quellcodeverzeichnis	IX
	Anhang	X

1 Machine Learning

1.1 Theorie

1.1.1 KI, ML, DL - Unterschiede

Im Zusammenhang mit maschinellem Lernen werden oft die Begrifflichkeiten "Künstliche Intelligenz", "Machine Learning" und "Deep Learning" verwechselt, jedoch handelt es sich bei diesen um eigene Bereiche, die sich in vielen Punkten überschneiden.

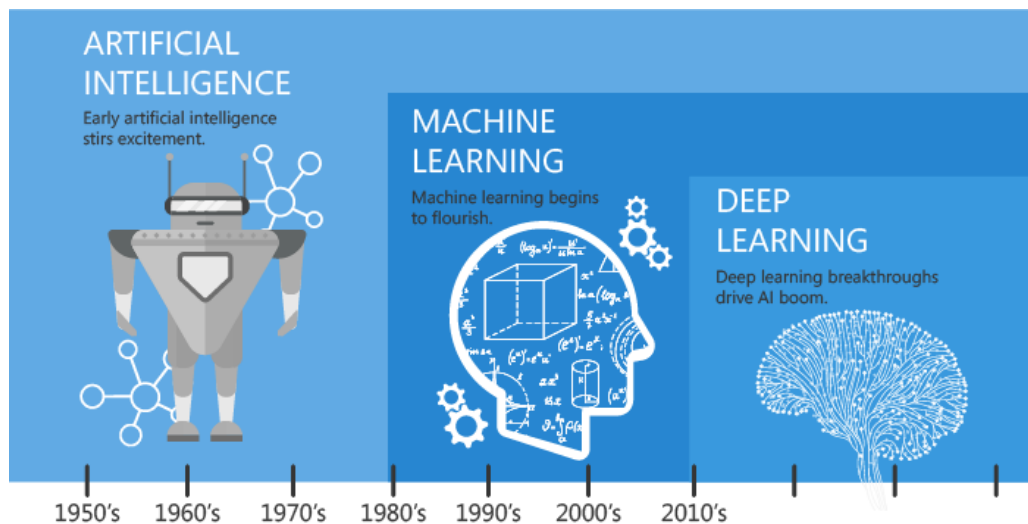


Abbildung 1: Evolution von Künstlicher Intelligenz

KI - Künstliche Intelligenz

Bereits in den 1950er erschien der Begriff "Künstliche Intelligenz", auf englisch "Artificial Intelligence", im Bereich Informatik, um Aufgaben zu lösen. Anfänglich noch mit dem Brettspiel "Dame" und einfachen Logikaufgaben. Jedoch kann es sich bei einer KI nur um eine programmierte Regel handeln, da man nur erwartet, dass sich die KI in gewissen Situationen auf eine bestimmte Art und Weise reagiert. Daher beschreibt es Programme, die humane Funktionen, imitieren und gibt nicht an mit welcher Technik das Problem gelöst wird.

ML - Machine Learning

Der Unterschied zur Künstlichen Intelligenz liegt beim Vorgang des Lernens. Genau wie ein menschliches Gehirn muss ein Machine Learning Model mit Daten trainiert werden, mit welchen das Model dann gewisse Klassifizierungen, Clusterbildungen oder Regressionen durchführen kann. Über die Zeit verbessert sich die Präzision, da diese mit Zuwachs der eingespielten Daten wächst.

DL - Deep Learning

Sowie beim Machine Learning sind Deep Learning Algorithmen abhängig von antrainierten Daten und kann daher als Synonym oder Untergruppe vom Machine Learning gesehen werden. Jedoch kann ein DL Model wie ein Mensch noch nie davor gesehene spezielle Bilder kategorisieren und ist damit einem ML-Modell in vielen Hinsichten überlegen.

Ohne Deep Learning würden die meisten modernen Assistenten nicht auf dem Niveau arbeiten wie erwartet, dabei handelt es sich bei Deep Learning um eine junge Technologie, die auf Neuronale Netzwerke basiert.

ML vs DL

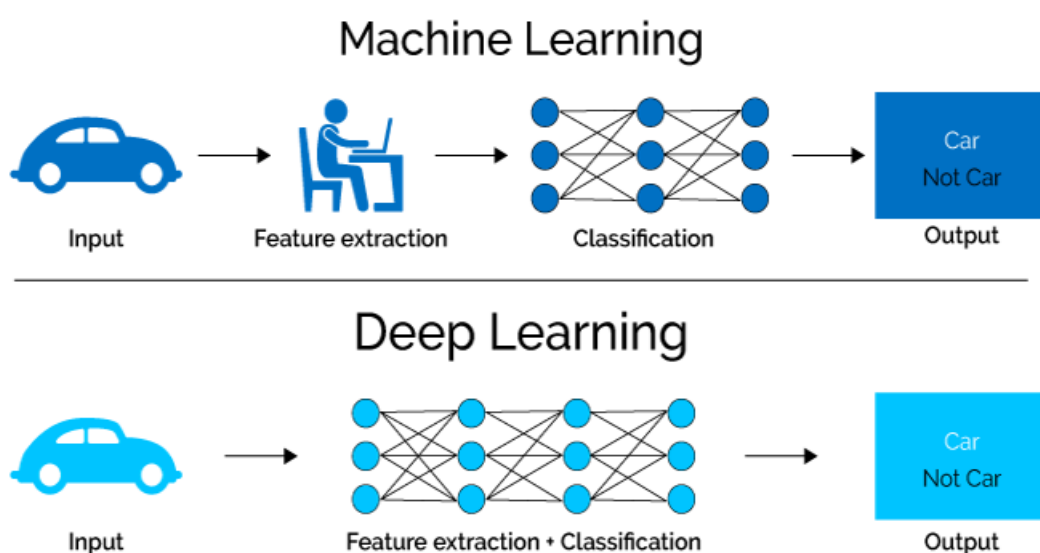


Abbildung 2: Feature Extraction, ML vs DL

Feature Extraction Bevor ein ML-Modell mit der eigentlichen Verarbeitung beginnen kann, müssen die Rohdaten einer sogenannten "Feature Extraction" unterlaufen,

welche die gegebenen Daten abstrakt darstellt. Dieser Prozess ist oft sehr kompliziert und beansprucht eine lange Zeit, außerdem ist hier eine Person notwendig, die sich in dem gegebenen Bereich auskennt.

Im Kontrast dazu gibt es Neuronale Netzwerke, welche den Feature Extraction Schritt übernehmen und selbstständig Rohdaten verarbeitet. Über mehrere Schichten werden bestimmte Merkmale hierarchisch definiert und später zum Beispiel zur Kategorisierung genutzt, hierbei erhöht sich die Genauigkeit der Extrahierung schon während des Antrainierens.

Welches Verfahren ist sinnvoller? In Bereichen, wo bereits strukturierte Daten vorhanden sind, bietet ein ML-Modell an, ein gutes Beispiel dafür ist die Erstellung von Vorhersehungen von Ereignissen. Außerdem benötigt man keine große Menge an Daten, um genaue Ergebnisse zu bekommen.

Stehen keine bereits vorpreparierte Daten zur Verfügung, wäre ein DL-Modell die bessere Lösung, aber dies auch nur wenn man eine große Menge an Daten hat.

1.1.2 Arten von Machine Learning

Der wichtigste und komplizierteste Teil beim maschinellen Lernen, ist die Kunst einem Computer das selbständige Lernen beizubringen, dabei orientiert man sich am Lernprozess eines Menschen. Dazu besteht es aus einer Vielzahl an Ansätzen und zu den bekanntesten gehören:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Um diese Ansätze nachvollziehen zu können, muss man zuerst die menschliche Intelligenz verstehen, oder genauer gesagt die Frage "Wie lernt das menschliche Gehirn?".

Menschliche Ursprünge vom Machine Learning

Ein Neugeborenes kommt mit circa 100 Milliarden Neuronen auf die Welt, diese sind jedoch nur schwach miteinander verknüpft. Mithilfe des Lernens werden diese Verbindungen gestärkt und das Kind kann Vorgehensweisen besser verstehen und neue Erkenntnisse gewinnen.

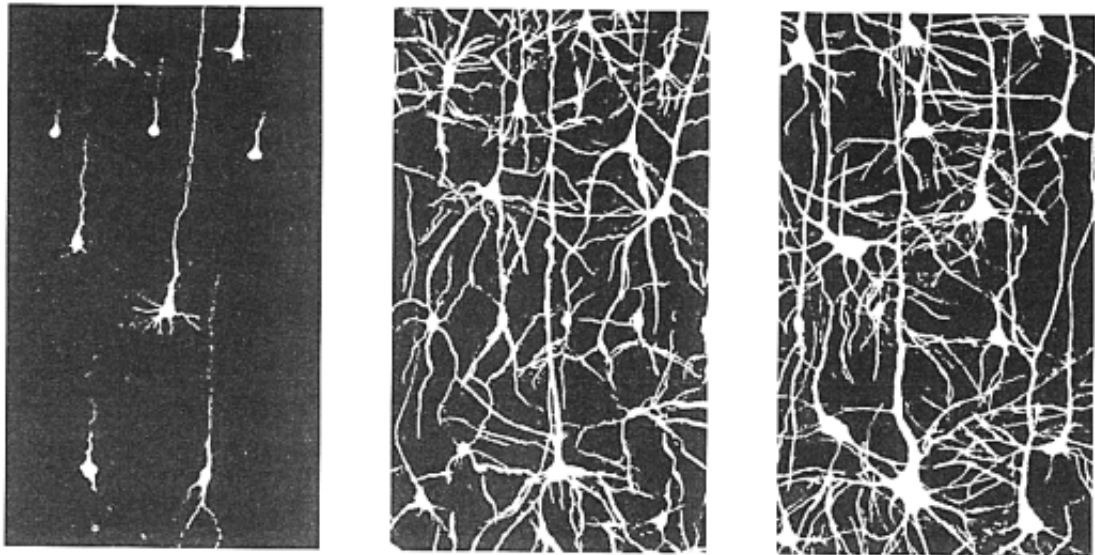


Abbildung 3: Vernetzungen nach der Geburt, nach 3 Monaten und nach 15 Monaten

Die gegenwärtige Neurobiologie, definiert das Lernen als vielschichtigen Prozess, welcher Neuronenverbände plastisch miteinander verbindet und so bilden sich Netzwerke und System aus Neuronen. Umso mehr Verbindungen entstehen und mit dem Wiederholen des Gelernten gestärkt werden, desto besser kann das Erlernte aufgerufen werden und mit dem bereits existierenden Vorwissen kombiniert werden.

Mithilfe einem zwecksentfermdeten, morphologischem Kasten, auch Zwicky-Box genannt, kann man das menschliche Gehirn vereinfacht darstellen. Dabei gibt die erste Spalte ausgewählte Merkmale an und die darauf folgenden Spalten eine weitere Ausprägung vom Merkmal in der selben Zeile. Die große Anzahl der Ausprägungen ermöglicht später breite Auswahl an Ausgängen der Zwicky-Box.

Merkmal	Ausprägungen		
Augenfarbe	Blau	Grün	Braun
Haarfarbe	Blond	Brünett	Schwarz
Frisur	Lang	Glaze	Kurz
Gesichtsform	Schmal	Rund	Oval
Nase	Klein	Groß	Flach
Brille	Ja	Nein	

Tabelle 1: Beispiel einer Zwicky-Box

In dieser Zwicky-Box wurden Merkmale eines Menschen aufgezählt. Um einen Ausgang darzustellen wird für jedes Merkmal eine Ausprägung ausgewählt und mit einer Linie

verbunden, damit man eine Ausprägung von anderen unterscheiden kann. In diesem Beispiel wird die Linie durch verschiedenen Farben ersetzt.

Merkmal	Ausprägungen		
Augenfarbe	Blau	Grün	Braun
Haarfarbe	Blond	Brünett	Schwarz
Frisur	Lang	Glaze	Kurz
Gesichtsform	Schmal	Oval	Rund
Brille	Ja	Nein	

Tabelle 2: Beispiel einer Ausprägung

Jede eingefärbte Zelle kann man als "aufleuchtendes" Neuron interpretieren, wenn man versucht das Geschlecht einer fremden Person zu identifizieren. Jedoch lernt man durch dieses "Aufleuchten" nicht, da man das eigentliche Geschlecht der Person nicht kennt und man daher keine Verbindungen aufbauen kann. Durch das Kennenlernen von Personen und Merken derer Ausprägungen werden diese oben genannten Neuronenverbindungen aufgebaut, welche später genutzt werden, um das Geschlecht einer fremden Person festzustellen. Diese Art von Lernen ist ein sehr schneller und effektiver Prozess, der mit vielen Nachteilen kommt, denn genau auf diesem Weg werden Vorurteile gebildet. Anhand eines Beispiels: Die meisten würden die obigen Werte aus der Zwicky-Box als Frau wahrnehmen, jedoch könnte jede kleinste Änderung am Aussehen diese Wahrnehmung beeinflussen.

Merkmal	Ausprägungen		
Augenfarbe	Blau	Grün	Braun
Haarfarbe	Blond	Brünett	Schwarz
Frisur	Lang	Glaze	Kurz
Gesichtsform	Schmal	Oval	Rund
Brille	Ja	Nein	
Ergebnis:	Frau		
Sicherheit:	70%		

Tabelle 3: Ergebnis einer Ausprägung

Da das Aussehen im Beispiel dem Aussehen von Bekannten in gewissen Merkmalen ähnelt, kann man mit einer Wahrscheinlichkeit von 70% sagen, dass es sich um eine Frau handelt. Desto mehr Personen man kennt, umso besser und genauer werden die Ausprägungen dem Geschlecht zugeteilt und die Sicherheit wird erhöht.

In dem Fall, dass ein Merkmal nicht erkennbar ist, wird versucht das beste Ergebnis zu finden, jedoch wird die Sicherheit drastisch fallen.

Supervised Learning

Hierbei werden gelabelte (daher auch "supervised" oder im deutsche überwacht) Datensätze genutzt, um Daten zu klassifizieren oder Vorhersagen genau wie möglich aufzustellen. Diese Art von Lernen kann man in zwei Typen aufteilen:

Klassifizierungs Probleme verwenden Algorithmen, um Daten einer bestimmten Kategorie zuzuteilen. Oft gibt es nur zwei Kategorien wie zum Beispiel Hund/Katze oder Ja/Nein, jedoch gibt es auch Fälle wo eine Vorhersage mit einer Wahrscheinlichkeit zwischen 0 und 1 getroffen wird. Weiteres gibt es auch Situationen, wo zwischen einer großen Menge an Kategorien ausgewählt wird, zum Beispiel bei der Erkennung von handschriftlichen Ziffern, in diesem Beispiel würde es zehn Möglichkeiten geben.

Zu diesen "überwachten" Algorithmen gehören Lineare Diskriminanzanalysen, Support Vector Machines (SVM), Random Forests und Entscheidungsbäume [1].

Entscheidungsbäume, oder im Englischem Decision Trees, sind Visualisierungen von Entscheidungswegen, oder auch aufgefächerte Zwicky-Boxen, die die Verbindungen visualisieren.

In diesem Beispiel sieht dieser Entscheidungsbaum noch sehr lesbar aus, jedoch ändert sich dies, wenn Entscheidungen dargestellt werden, bei denen es auf die Nachkommastelle ankommt und wenn die Kategorien sehr schwer differenzierbar sind. Bei dem Supervised Learning erstellt das Programm selbstständig einen Entscheidungsbaum, indem es Muster oder Zusammenhänge findet und analysiert. Nach vielem Lernen kann dieser Entscheidungsbaum optimiert werden und unnötige Verbindungen können entfernt werden. Jedoch kann es auch passieren, dass durch Zufälle unberechenbare Verbindungen gefunden werden, dieses Problem wird "Overfitting".

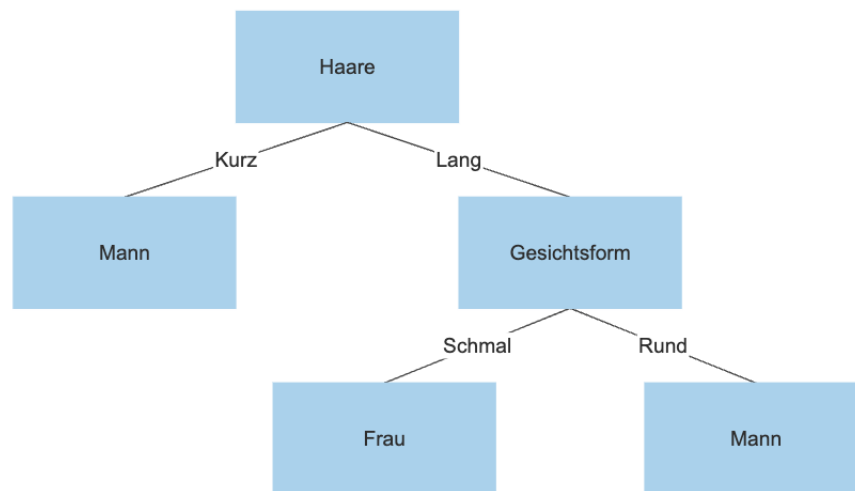


Abbildung 4: Decision Tree an dem Beispiel von 2

Overfitting bedeutet, dass ein Modell perfekt an die Testdaten angepasst ist, jedoch bei unbekannten Daten falsche Verbindungen aufbauen könnte und dadurch falsche Ergebnisse vorhergesagt werden.

Regressionen, Erstellung einer kontinuierlichen Funktion mithilfe von Werten, die auf oder nahe an der Funktion liegen, sind hilfreich, wenn anstatt diskreten Werten kontinuierliche Werte, wie zum Beispiel die Größe einer Person, festgestellt werden sollen. Dazu gehören lineare Regressionen und logistische Regressionen.

Unsupervised Learning

Beim Unsupervised Learning, oder Unüberwachtes Lernen, erzeugt Verbindungen ohne genauere Informationen über den Testdatensatz. Dabei muss das Programm selbst Gruppen definieren und dann die übergebenen Daten in diese Gruppen zuordnen.

Clustering ist einer der beliebtesten Variante, um selbstständig Gruppen zu erstellen. Dabei wird jeder Datensatz als Punkt in ein Koordinatensystem mit beliebig vielen Dimensionen eingetragen. Eine Achse stellt ein Attribut dar und je nach Ausprägung ist der Punkt mehr oder weniger vom Ursprung entfernt.

Auch für das menschliche Auge ist es möglich dieses Beispiel 5 in Haufen oder Klumpen zusammenzufassen, genau das gleiche macht ein Programm mit Clustern. Die Interpretation dieser Gruppen muss jedoch wieder durch Menschen erfolgen, da ein Computer nicht im Stande dazu ist, dieser Cluster einer Kategorie zuzuteilen.

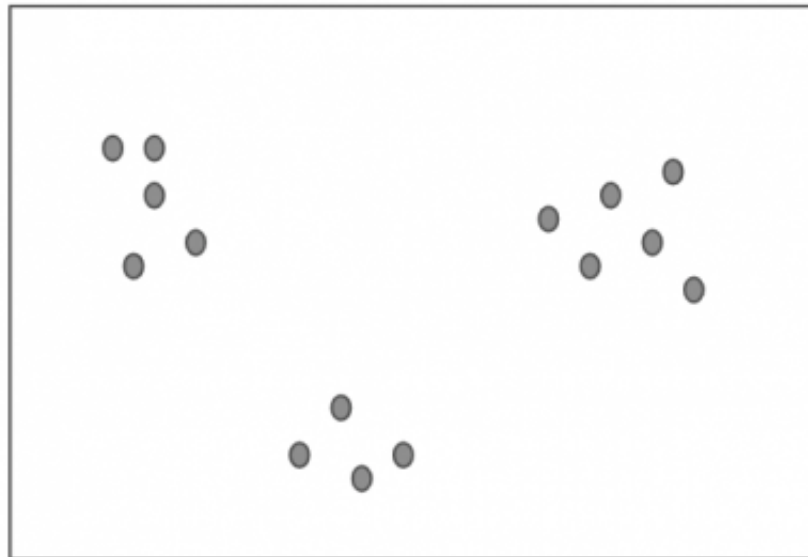


Abbildung 5: Unkategorisierte Daten

Diese Vorgehensweise wird oft in sehr komplizierte Einsatzbereiche genutzt, und daher ist es oft sehr schwer differenzierbare Cluster zu erstellen. Der Prozess, solche Cluster zu definieren, basiert darauf die Punkte so zu gruppieren, dass der Abstand in diesem Cluster klein ist und zu anderen Clustern groß.

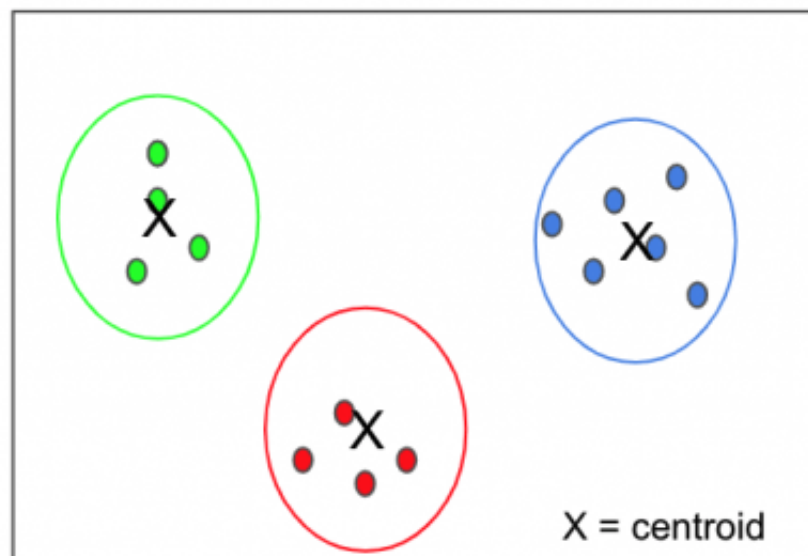


Abbildung 6: Kategorisierte Daten

Reinforcement Learning

Das Reinforcement Learning oder das beschränkte/verstärkte Lernen wird oft mit dem Konzept "Learning by doing" verglichen, da es sich weniger auf das Ergebnis fokussiert und mehr auf Aktionen oder Vorgängen. Ein Beispiel für diese Vorgehensweise aus der

Sicht eines Schülers ist das Üben vor einer Matheschularbeit. Wird während dem Üben ein Fehler gemacht, merkt man sich das Problem und passt sein Verhalten / seinen Rechenweg so an, dass dieser Fehler nicht mehr vorkommt. Die nennt man auch negative Verstärkung.

Dahingegen führen richtige Ergebnisse oder erwarteten Reaktionen zu positiven Verstärkungen und man versucht dieses Verhalten zu wiederholen.

Durch negative und positive Verstärkungen wird das Verhalten verbessert, um den besten Weg zum Ziel zu finden. Bei komplexen Systemen kann dies selbstständig vom Programm gemacht werden, jedoch bei simpleren kann es geschehen, dass ein unnötig komplizierter Weg gefunden wird. In diesen Fällen schaut ein Supervisor dem Programm "über die Schulter".

1.1.3 Optical Character Recognition

In den meisten Fällen erfolgt eine Eingabe für eine Tastatur, jedoch ist dies manchmal weder die schnellste noch effizienteste Art, Text einzulesen. Mithilfe von Optical Character Recognition ist ein automatisiertes Einlesen und Verarbeiten möglich und das schon bereits in den 1950er. Am Anfang noch um Verkaufsberichte in Lochkarten zu konvertieren, damit ein Computer mit den Verkaufsdaten arbeiten kann [2].

Im Bereich von OCR sind bereits jetzt gute und genaue Resultate mit Machine Learning erwartbar, jedoch wie bei allen anderen Problem ist es verbesserbar. Um einen großen Fortschritt zu erreichen, würde die Nutzung von Deep Learning unausschließbar sein, dies ist jedoch in den meisten Situationen nicht notwendig.

Die Präzision hängt von vielen Attributen ab, dabei kann ein eingescannter Text viel besser verarbeitet werden als ein in der Freien geschossenes Bild mit dem Fokus auf ein Straßenzeichen [3].

- Textdichte

Es macht einen Unterschied wie viel Text sich auf einer Fläche oder einem Bild befindet, denn es ist in gewissen Situationen leichter Text auszulesen, wenn dieser nur spärlich vorkommt.

- Struktur

Wenn man eine klare Struktur erkennt, zum Beispiel in Tabellen oder in Zeilen, kann man ein besseres Ergebnis erwarten, daher ist es auch wichtig, dass man

vor dem Auslese-Prozess das übergebene Bild aufbereitet und als Beispiel die Rotation ändert.

- Schriftart

Handgeschriebene Texte oder "laute" Schriftarten sind im Gegensatz zu einfachen und gedruckten viel komplizierter, da sie kaum strukturiert sind.

- Buchstaben

Sprachen wie Arabisch, Chinesisch, Russisch oder Japanisch benutzen im Gegensatz zu Deutsch ein anderes Alphabet, dabei kann es zu ähnlichen Buchstaben und Vertauschungen kommen, daher sinkt die Präzision in Texten mit mehreren Sprachen. Dies kann auch der Fall sein, wenn mathematische Formeln vorkommen.

Lateinisches Alphabet	Kyrillisches Alphabet	Ähnelt dem Buchstaben
r	p	p
B	B	B
I	H	H
U	y	y
S	C	C

Tabelle 4: Ähnlichkeiten zwischen Buchstaben im Lateinischem und Kyrillischem Alphabet

- Platzierung

Zentrierte Texte erlauben ein besseres Auslesen, als abgeschnittene oder verstreute Wörter.

Strategien

Die Texterkennung ist generell in zwei Phasen aufgeteilt:

Text detection ist der Prozess, indem in einem Bild oder einer PDF erkannt wird, wo sich Text befindet. Beim Resultat handelt es sich um Bounding Boxen, diese schließen einen Textblock (Wort, Buchstabe oder Paragraph) ein, dabei werden die genauen Koordinaten der Eckpunkte mitzurückgegeben. Diese werden später genutzt, um zur Visualisierung Boxen auf dem Bild oder der PDF aufzuzeichnen. Dieser Prozess wird auch in der Objekterkennung genutzt.

Qty	Rate	Current
40.00	16.50	660.00
		Current
		0.00
		-84.00

Abbildung 7: Bouning Boxes Beispiel

Die Erkennung kann entweder mittels der auf Regionen basierenden oder Textur basierenden Methode durchgeführt werden.

Bei der **Regions basierenden** Methode, werden Pixel verbunden und als Zeichenkandidat markiert, welche später mehrmals gruppiert werden und schlussendlich Wörter oder Textzeilen bilden. Dabei kommt es auf die geometrischen Eigenschaften, dabei kann es jedoch zu Segmentierungs Fehlern kommen. Wie man bei 7 sehen kann, wurde der Buchstabe "C" im Wort "Custom" oder die letzte Nachkommastelle bei "40.00" nicht ganz als Teil des Wortes oder der Zahl erkannt.

Mit dem Stroke Width Transformer (SWT) wird jedem Pixel eine Strichbreite zugeteilt, indem zwei Kanten gefunden werden mit der gleichen Richtung modulo 180° . Die Entfernung dieser zwei Kanten werden in den Kantenpixeln und alle unterliegenden Pixeln als Strichbreite gespeichert und alle Zusammenliegenden, gleich breite Pixel werden zu einem Zeichenkanidaten gruppiert. Danach werden alle benachbarten Zeichenkanidaten untersucht und zu einem Wort gruppiert, falls das mittelwertige Strichbreite Verhältnis nicht über 2 liegt. Außerdem wird die Höhe und Farbe des Zeichenkanidaten berücksichtigt. Dies kann auch der Grund sein, wieso bei 7 das Minuszeichen bei "-84.00" nicht zur Zahl hinzugefügt wird, das Minuszeichen ist signifikant niedriger als der Rest der Zahl. [4]

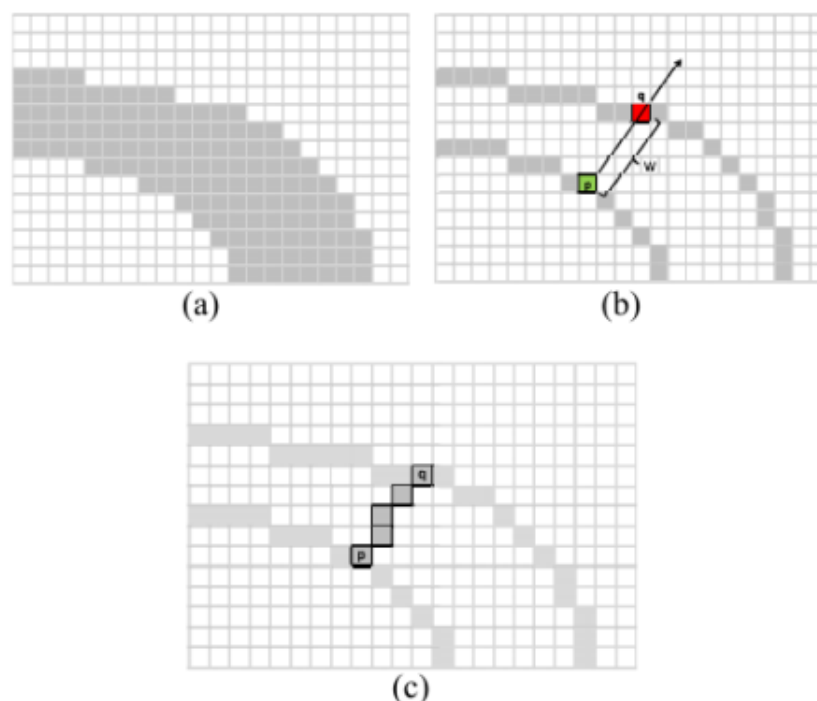


Abbildung 8: Die Kanten des Striches (a) werden solange verglichen, bis zwei gefunden werden mit der mit der gleichen Richtung (b). Alle unterliegenden Pixel erhalten die Strichbreite der Entfernung zwischen der Start- und Endkante (c).

Textur basierende Methoden unterteilen das Bild in Fester, dessen Höhe wird später mit der geschätzten Textgröße verglichen. Dabei kann es ebenfalls zu Erkennungsfehlern kommen.

Die Mischung dieser beiden Methoden gewann im Jahr 2011 den ICDAR-Wettbewerb 1.1.3 mit einem F-Score von 71.28%. Hierbei hat Chunghoon Kim den Vorschlag gegeben, als erstes werden Blöcke extrahiert mit dem Maximally Stable Extremal Regions (MSER) Verfahren, danach werden benachbarte Blöcke gruppiert, falls die Farbe und Größe sich ähnelt. Jedoch werden mit diesem Verfahren ebenfalls eine große Menge an false-positive Blöcken erkannt. Um diese Anzahl zu verringern wird eine ähnlich Idee zur SWT verwendet.

Wettbewerbe sind ein großer Grund für die Fortschritte in der Texterkennung und im Rahmen des zwei-jährlich stattfindenden International Conference of Document Analysis and Recognition (ICDAR) Wettbewerbs wurden alle oben genannten Ideen verglichen [5].

Text recognition ist genau wie bei der Erkennung von Text in zwei Möglichkeiten unterteilt: Regions basierend und Textur basierend.

Die von MSER generierten und normalisierten Blöcke werden je nach der Orientierung in ein separates Bild extrahiert. Dabei sind acht Orientierungen möglich, welche mit einem Gaußschen Filter bearbeitet werden und auf ein 5×5 Bild komprimiert wurden. Mit diesen $5 \times 5 \times 8 = 200$ dimensional Vektoren werden, dann die bereits erkannten Blöcke klassifiziert.

2 Cloud Computing

2.1 Theorie

2.1.1 Was ist Cloud Computing?

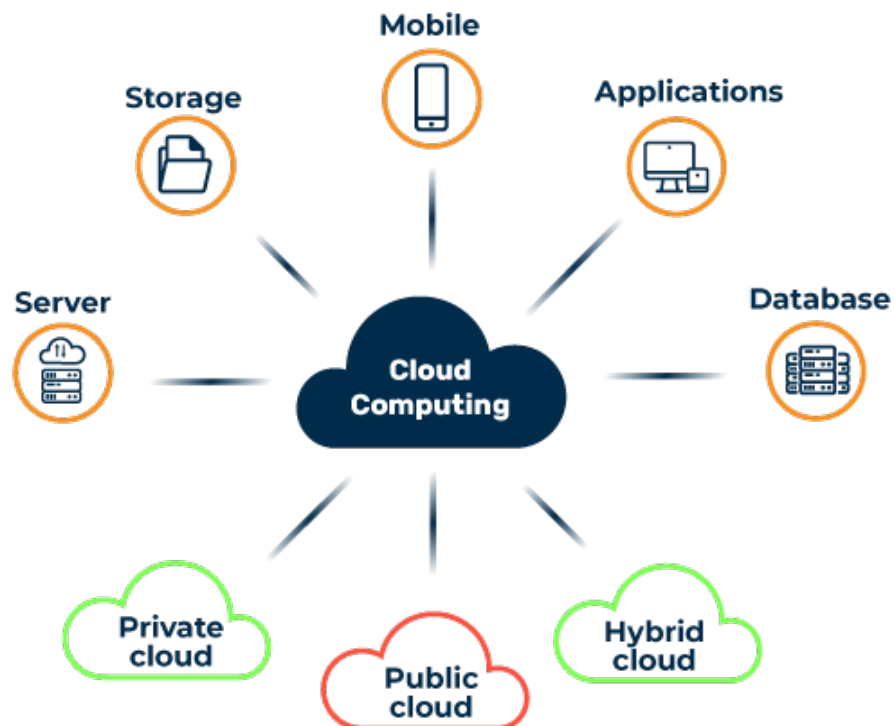


Abbildung 9: CC-Overview

Cloud Computing ist ein Modell, bei dem virtueller Speicher, jegliche Art von Server, Anwendungen nicht physisch, sondern rein über das Internet bereitgestellt werden. Diese werden in der Regel nach Bedarf als Teil eines As-a-Service-Modells angeboten. Mithilfe der Cloud werden Organisationsinterne Systeme und Rechenzentren mit virtuellen Ressourcen, wie Datenbanken und spezielle KI- und Rechensysteme ausgetauscht. Im Normalfall werden diese Ressourcen von externen Anbietern bereitgestellt und gewartet.

2.1.2 Vorteile von Cloud Computing

Oft fallen, in Verbindung mit dem Thema Cloud Computing, Stichwörter wie „flexibel“ oder „agil“. Das liegt daran, dass gerade in heutigen Zeiten marktseitige und auch technologische Veränderungen schnell und oft vorgenommen werden. Deshalb betreiben viele Unternehmen sogenanntes Outsourcing, um nicht selbst für die Rechenleistung ihrer eigenen Dienste verantwortlich zu sein. Skalierbarkeit spielt hierbei eine große Rolle, ein Beispiel hierfür: je mehr Zugriffe auf einen Webshop erfolgen, desto mehr Ressourcen müssen im Hintergrund hochgefahren werden, um eine fehlerfrei Nutzung zu garantieren.

Auch die Usability ist simpler bei cloudbasierten Prozessen, da viele dieser Prozesse in den Hintergrund verschoben und somit von den Cloud Service-Anbietern übernommen werden. Der Aufwand für die Wartung und Beschaffung für Rechenzentren entfällt weitgehend. Demzufolge kann man im Bereich Energie- und Erhaltungskosten einsparen. Im Allgemeinen werden die Kosten für ebensolche Dienste je nach Absprache nutzungsabhängig festgelegt. Im Normalfall fallen diese Kosten monatlich oder jährlich an, wobei diese verhältnismäßig kleiner als die der On-Premise Lösungen sind.

Ein essentieller Bestandteil ist die Datenkonsistenz. Im Fall von komplexen Prozessen ist die Konsistenz der Daten ein wesentlicher Punkt, um drastische Probleme zu verhindern. Gerade bei dezentraler Speicherung und Verarbeitung der Daten, ist die Synchronität überaus wichtig. Bei Cloud Computing ist dieses Risiko minimiert. Der Grund hierfür ist, dass die Daten, auch bei einem Zugriff von unterschiedlichen Schnittstellen synchron ist.

2.1.3 Hindernisse für den Einsatz einer Cloud

Da es sich beim Cloud Computing um ein neues Modell handelt, besteht eine gewisse Unsicherheit, inwiefern es möglich ist, auf allen verschiedenen Ebenen eine solide Sicherheit zu erreichen. Dadurch wird die Fähigkeit der Cloud, den Datenschutzbestimmungen gerecht zu werden, in Frage gestellt. Das Grundprinzip der Cloud sieht eine dauerhafte Verfügbarkeit vor, jedoch kann gerade diese Verfügbarkeit auch ein großer Nachteil sein. Erst durch den offenen Zugang zu den zur Verfügungen gestellten Rechenleistung und anderen Ressourcen, entfaltet das CC-Model sein volles Potenzial.

Heutzutage müssen Anwendungen dauerhaft erreichbar sein. Hierbei kann sich die Abhängigkeit von einem Cloud Service Provider negativ auf die dauerhafte Konnektivität

auswirken. Im Fall von Aussetzern oder kompletten Ausfällen müssen Notfallpläne oder neue Ressourcen gestartet werden, um Datenverlust zu verhindern.

2.1.4 Cloud Computing Modelle

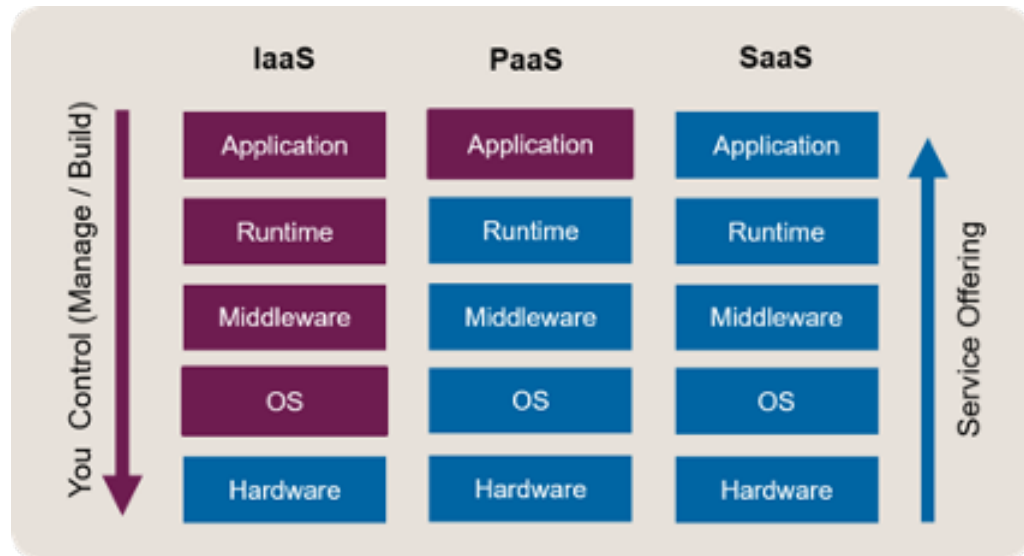


Abbildung 10: Modelle

Infrastructure as a Service (IaaS)

Infrastructure as a Service stellt sowohl virtuelle und physische Server, als auch Netzwerk- und Speicherressourcen, nach Verlangen des Verbrauchers, zur Verfügung. Der Nutzer hat die völlige Kontrolle über Ressourcen, in Bezug auf Speicher und Rechenleistung. Die Auswahl des Betriebssystems wird ebenfalls dem Nutzer überlassen. Beispiele hierfür sind Amazon EC2-Cluster und Microsoft Azure.

Platform as a Service (PaaS)

Diese Form des Cloud Computing gibt dem Kunden die Möglichkeit, seine Anwendungen auf einer, vom Dienstleistungsgeber, gehosteten Plattform zu entwickeln, bereitzustellen und zu verwalten. Bei diesem Modell handhabt der Dienstleister sämtliche Ressourcen. Die Anwendung wird den potenziellen Nutzern über APIs zur Verfügung gestellt. Hierbei ist wichtig, dass der Nutzer keinerlei Einfluss über, die im Hintergrund laufenden, Ressourcen hat. Bekannte Beispiele hierfür sind Web-Hosting Dienste, wie etwa +Microsoft Azure Web und Amazon Web Services.

Software as a Service (SaaS)

Bei SaaS wird den Nutzern die Möglichkeit gegeben auf eine, vom Dienstleister in der Cloud Infrastruktur bereitgestellten Anwendung, zu zugreifen und zu benutzen. Benutzer können dann über eine webbasierte Schnittstelle, wie beispielsweise ftp und clientseitige Schnittstellen, darauf zugreifen. Jene Anwendungen werden gegen monatliche oder jährliche Zahlungen dem Benutzer bereitgestellt. Jedoch hat der Verbraucher keine bis wenig Kontrolle über im Hintergrund geregelte Ressourcen. Beispiele hierfür sind Microsoft Office 365, Microsoft Skype und Google Apps.

2.1.5 Anwendungsfälle von Cloud Computing

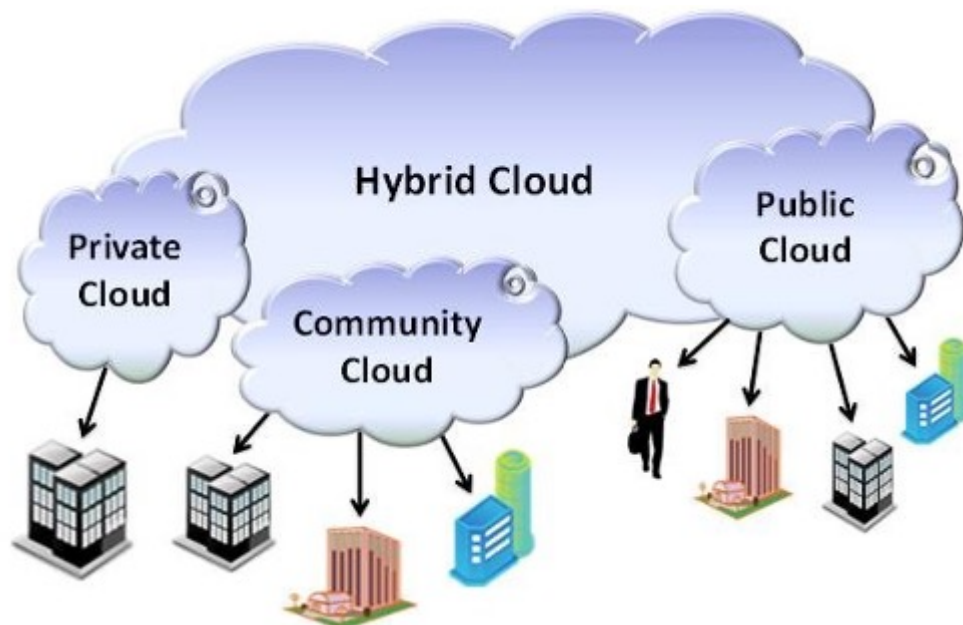


Abbildung 11: Anwendungsfälle

Private Cloud

Dieses Bereitstellungsmodell wird vor allem in Unternehmen verwendet, das heißt diese Dienste sind nicht für die Öffentlichkeit zugänglich. Sie werden nur Unternehmensintern verwendet. Die Nutzer können standortunabhängig von diesen Cloud Computing Diensten Gebrauch machen, müssen aber Teil der gleichen Organisation sein. Die private Cloud ist die sicherste der Bereitstellungsmethoden. Der Grund dafür ist, dass Prozesse innerhalb des Unternehmens kontrolliert und verwaltet werden, ohne jegliche Leistungs-Sicherheitsbeschränkungen, die Dienste in der Public Cloud vielleicht benötigen. In der privaten Cloud ist es möglich, dass die fundamentale Infrastruktur der Cloud vom

Unternehmen selbst, von Drittanbietern oder auch von Beiden verwaltet wird. Generell wird die Private Cloud in zwei unterschiedliche Arten unterteilt:

- On-Premise Private Cloud

Diese Art wird auch interne Cloud genannt. Sie bietet zusätzliche Sicherheit, dennoch kann sie in Größe und Skalierbarkeit stark eingeschränkt sein, da man selbst das Kapital für Hard- und Software, sowie Wartung und Instandhaltung aufbringen muss. Die interne Cloud eignet sich somit für Anwendungen, die die volle Kontrolle sämtlicher Ressourcen verlangen.

- extern gehostete Private Cloud

Hier wird das Hosten der Cloud von einem Drittanbieter übernommen. Diese Anbieter fördern eine restriktive Cloud Computing-Umgebung mit vollständiger Vertraulichkeit. Diese Art der Private Cloud wird für Institutionen empfohlen, die das Kapital für eine interne Private Cloud nicht aufbringen können.

Public Cloud

Eine Public Cloud ist für die Öffentlichkeit zugänglich und kann von einem Unternehmen, einer staatlichen Einrichtung oder Organisation verwendet werden. Die Cloud ist jedoch im Besitz eines Drittanbieters. Hierbei greift der Nutzer über Schnittstellen auf die CC-Dienste des Cloud-Besitzers zu.

Community Cloud

In der Community Cloud wird die Infrastruktur einer Cloud von mehreren Unternehmen oder Institutionen mit ähnlichen oder gemeinsamen Zielen verwendet. Die Infrastruktur kann entweder von einer Organisation oder mehreren Organisationen verwaltet werden.

Hybrid-Cloud

Diese Variante ist eine Kombination aus einer oder mehreren privaten und öffentlichen Cloud, die aber als getrennte Einheiten fungieren. Diese Einheiten werden durch eine Protokollen und Standardisierungen verbunden. Dieses Modell wird hauptsächlich verwendet, wenn man die Vorteile der unterschiedlichen CC-Bereitstellungsoptionen kombinieren möchte. Beispielsweise, ein Unternehmen möchte die Datenspeicherung

auf einer privaten Cloud realisieren, und andere Aufgaben mithilfe einer öffentlichen Cloud erledigen.

Azure Cloud

Microsoft Azure ist ein Cloud-Computing-Dienst von Microsoft. Azure bietet eine Reihe von Software as a Service (SaaS), Plattform as a Service (PaaS) und Infrastruktur as a Service (IaaS) Optionen, für die Bereitstellung von Anwendungen und Diensten, auf einer von Microsoft verwalteten Rechenzentrumsinfrastruktur an. Mit 50 Betriebsregionen bietet Azure mehr als jeder andere Cloud-Anbieter.

2.2 Praxis: AI Builder

Basierend auf Azure AI Cognitive Service ist AI Builder ein Tool, welches zum Erstellen und Trainieren von Modellen, ohne das Schreiben von Code dient. Die Integration mit Power Apps und Power Automate ist eine Funktion, die den Nutzern die Möglichkeit bietet, bestehende Geschäftsanwendungen zu erweitern, und zu verbessern. Microsoft Power Plattform ist eine Low-Code-Plattform, die es Unternehmen bewilligt, Geschäftsprozesse zu automatisieren. Power Plattform umfasst drei Hauptprodukte: Power BI, PowerApps und Flow. Mit dem AI Builder ist es möglich, auf einfache Weise Prozesse zu automatisieren und Ergebnisse vorherzusagen. AI Builder ist eine schlüsselfertige Lösung, die die Leistungsfähigkeit der künstlichen Intelligenz von Microsoft mit wenigen Mausklicks anwendbar macht. Mit dem AI Builder kann man Anwendungen Intelligenz beifügen, auch wenn Sie keine Programmier- oder Data-Science-Kenntnisse haben.

2.2.1 Benutzerdefinierte Modelle

Der erste Schritt bei der Erstellung eines KI-Modells besteht darin, festzustellen, ob für den Anwendungsfall bereits vorgefertigte oder bereits trainierte Modell vorhanden sind. Ist dies nicht der Fall, stehen in der Benutzeroberfläche des AI Builders fünf Modelle zur Verfügung:

1. Category Classification
2. Entity Extraction
3. Form Processing

4. Object Detection

5. Prediction

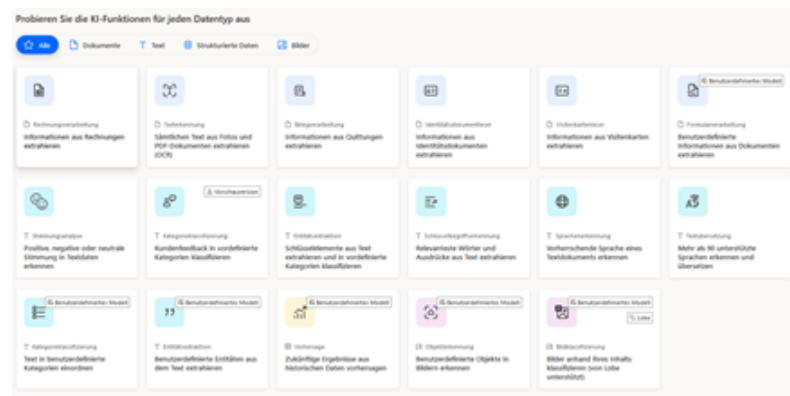


Abbildung 12: AI-BUILDER Modelle

Category Classification

Bei diesem Ansatz wird ein Modell verwendet beziehungsweise trainiert, um große Mengen an Textdaten, Dokumenten oder sonstige Textdatenquellen zu analysieren, und den Text zu klassifizieren. Besonders hilfreich ist dieses Modell, um Spam zu identifizieren, und entsprechend zu behandeln. Zuerst muss das Modell mit Trainingsdaten trainiert werden. Das Zeichenlimit für jede Textprobe liegt bei fünftausend Zeichen.

Die Analysen, die dieses Modell liefert, können auch als Input für andere KI-Lösungen verwendet werden. Wichtig hierbei ist, dass für jeden Tag mindestens zehn Textproben bereitgestellt werden, ansonsten sinkt die Wahrscheinlichkeit ein genaues Ergebnis zu erzielen.

Entity Extraction

Hier werden wichtige Textelemente identifiziert, ferner den definierten Kategorien zugeordnet. Die Ergebnisse werden dabei, entsprechend den Anforderungen, standardisiert und strukturiert. Auch hier werden wieder mindestens 10 Datensätze benötigt, um mit dem Trainieren des Modells beginnen zu können. Das Modell ist anpassbar, indem man neue Entitätstypen mit wenigen Trainingsdaten erstellt oder bestehende Entitätstypen modifiziert. Der Ai Builder verfügt über vorgefertigte Trainingsdaten, die zur Erweiterung der eigenen Trainingsdaten verwendet werden können.

Form Processing

Die Formularverarbeitung ist das KI-Modell, das Daten aus Formularen, auch aus Papier- oder PDF-Dokumenten, extrahiert. Es werden fünf Beispielformulare benötigt, um es zu trainieren, die Felder eines Dokuments zuzuordnen und eine funktionierende Anwendung zu erstellen. Diese Lösung wird verwendet um Rechnungen und Aufträge zu erfassen. Beispielsweise ist es möglich, das Modell zu trainieren und einen Ablauf zu erstellen, der automatisch Schlüsselinformationen aus Bestelldokumenten erkennt, extrahiert und anschließend eine E-Mail an den zuständigen Mitarbeiter sendet.

Die empfohlenen Formate für die Eingabedaten sind .jpg, .png und .pdf. Die Gesamtgröße der, für das Training verwendeten, Dokumente, darf insgesamt 50 MB nicht überschreiten.

Object Detection

Die Objekterkennung wird verwendet, um Objekte auf Fotos oder Videos zu erkennen. Dieses Modell kann verwendet werden, um bestimmte Informationen von Produkten oder Maschinen zu erhalten. Ebenfalls hilfreich kann dieses Modell bei mobilen Anwendungen sein.

Für das Training werden mindestens 15 Fotos von jedem Objekt benötigt; je mehr Fotos, desto genauer ist das Modell. Um eine korrekte Identifizierung zu gewährleisten, sollten die Fotos möglichst verschiedene Hintergründe beinhalten. Ebenfalls empfehlenswert ist es, Fotos von Objekten aus diversen Entfernungen und Winkeln bereitzustellen. Es ist zu beachten, dass die Trainingsbilder im .jpg-, .bmp- oder png-Format vorliegen müssen und insgesamt 6 MB pro Training nicht überschreiten dürfen. Die Grenze von 256 x 256 Pixel soll jedoch nicht unterschritten werden.

Prediction

Bei diesem Modell werden große Mengen an alten Daten analysiert, um darin Muster zu erkennen. Dieses „Wissen“ wird dann verwendet, um diese Muster in neuen Datensätzen zu erkennen und Vorhersagen zu treffen. Diese Mechanismen können Muster aufdecken, die binäre Fragen (ja/nein), Fragen mit mehreren Antworten (eine Liste von Ergebnissen) oder Fragen, die mit einer Zahl beantwortet werden.

Zum Trainieren des Modells werden mindestens 10 Zeilen mit historischen Werten für jede Klasse der Datenspalte "Label" benötigt. Die Mindestanzahl der Zeilen für

das Training beträgt 50, aber ein Minimum von 1.000 Zeilen gewährleistet die besten Ergebnisse.

2.2.2 AI-Builder in der Praxis

Um den AI-Builder für den gewünschten Use-Case am besten zu nutzen, hat sich das Diplomarbeitsteam dazu entschieden, das Modell des Form Processing zu verwenden. Der gewünschte Use-Case ist, mehrere im Vorhinein definierte Felder aus einer Eingangsrechnung, im PDF, zu extrahieren und anschließend zu markieren. Folgende Schritte sind notwendig, um ein Modell zu erstellen, dass auf spezielle Eingangsrechnungen (Siehe Abbildung 13) trainiert ist:

1. Zu extrahierende Felder definieren
2. Trainingsdaten bereitstellen
3. Trainingsdaten mit Tags versehen
4. Modell trainieren und
5. Modell veröffentlichen

Musterfirma A
4020 Musterstrasse a
Musterstadt A
Musterland A

RECHNUNG
12345

Rechnung an:
Musterfirma B
4020 Musterstrasse b
Musterstadt B
Musterland B

Datum: Oct 21, 2020
Fälligkeitsdatum: Nov 11, 2020

Saldo fällig: 1.270,00 €

Artikel	Menge	Rate	Betrag
Musterprodukt A	3	400,00 €	1.200,00 €
Musterprodukt B	10	7,00 €	70,00 €

Zwischensumme: 1.270,00 €
MwSt (0%): 0,00 €
Gesamt: 1.270,00 €

Abbildung 13: Beispiel Eingangsrechnung

Zunächst ist es wichtig zu wissen, welche Felder man aus der ER extrahieren will. In der folgenden Abbildung wird gezeigt welche Felder bei dieser Arbeit ausgewählt

wurden. Diese Felder wurden nur zu Veranschaulichung der Fähigkeiten des AI-Builders ausgewählt und nicht zur firmeninternen Nutzung:

Zu extrahierende Informationen auswählen

Listet alle Informationen auf, die vom KI-Modell aus den Dokumenten extrahiert werden sollen. Beispiel: Name, Adresse, Gesamtbetrag. Sie markieren diese in den Dokumenten.

+ Hinzufügen ▾







 Name	Typ
 Company Name	Feld
 Company Address	Feld
 Invoice Id	Feld
 Invoice Date	Feld
 Total Value	Feld
 Items	Mehrseitige Tabelle (Description, Qty, Unit, Unit Price, Total Price)

Abbildung 14: AI-Builder Felder

1. Company Name
2. Company Address
3. Invoice Id
4. Invoice Date
5. Total Value
6. Items (Mehrseitige Tabelle)

Items (Mehrseitige Tabelle): Diese Tabelle bietet ein experimentelles Feature, um tabellarische Daten aus einer, in der Rechnung vorhandenen Tabelle zu entnehmen, auch wenn sich die Tabelle über mehrere Seiten zieht. Diese Tabelle enthält wiederum eigene Daten wie:

1. Description
2. Qty (Quantity)
3. Unit
4. Unit Price
5. Total Price

Um Daten für das Trainieren des Modells bereitzustellen, ist es notwendig mindestens fünf Beispieldokumente hochzuladen. Nichtsdestotrotz ist für eine hohe Genauigkeit des Modells, sprich wie sicher sich der Algorithmus ist, dass er das richtige Feld markiert hat, von Vorteil mehrere ähnliche Dokumente bereitzustellen. Um dem Algorithmus

anzutrainieren, welche Felder in dem Dokument vorhanden sind, ist es relevant die Daten in dem Dokument händisch zu markieren und die zuvor definierten Informationsfelder zu zuweisen. Siehe folgende Abbildung (20)

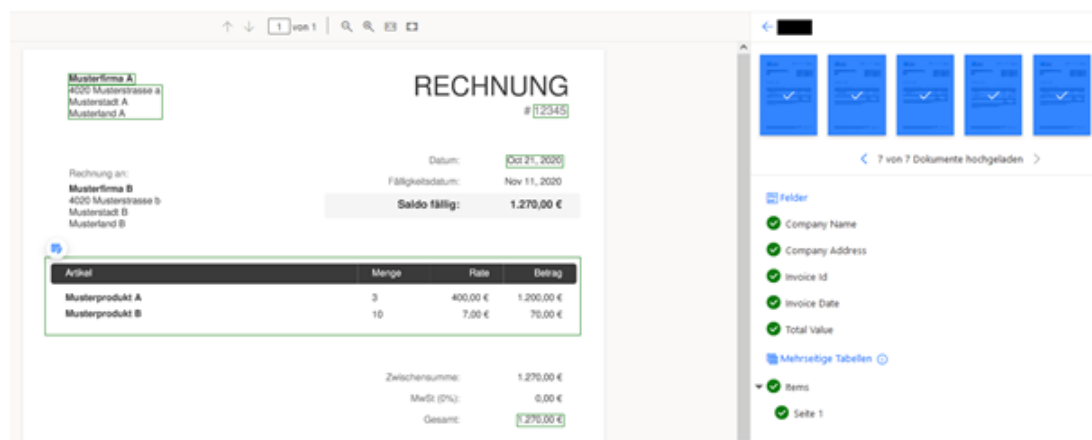


Abbildung 15: AI-Builder Felder Tagging

2.3 Praxis: Power Automate

Power Automate ist ein Teil der Microsoft Power-Plattform, mit der man Cloud-basierte Automatisierungsprozesse erstellen und verwalten kann. Der Benutzer hat Zugriff auf verschiedene Dienste von Microsoft und kann ohne oder wenig Code einen Workflow erstellen. Power Automate reagiert auf einen im Vorhinein definierten Trigger und führt eine Reihe von vordefinierten Schritten aus. Dieser Trigger kann eine neue E-Mail, ein neuer Datensatz im CRM eingefügt wird oder einfach die Erstellung eines neuen Dokuments im SharePoint sein. Somit ist es möglich auch ohne jegliche Vorkenntnisse oder fachspezifisches Wissen eine einfache Anwendung zu erstellen. Zum Beispiel kann beim Eingang einer E-Mail, die eine Kundenrechnung enthält, an den AI-Builder weitergegeben werden und dort mit einem vortrainierten Modell, die wichtigsten Daten herauslesen und dann erneut mit einer E-Mail versendet werden.

2.3.1 Cloud Flow

Da bei dieser Arbeit nur der Cloud Flow in Verwendung ist, wird hier nur auf diesen genauer eingegangen. Power Automate bietet hier drei verschiedene Arten von Cloud Flows an:

1. Automated Flow
2. Instant Flow

3. Scheduled Flow

Automated Flow

Dieser Flow wird ausgelöst, wenn eine gewünschte Kondition eintritt. Solche Trigger können, der Eingang einer E-Mail einer speziellen Person, eine Teams-Nachricht oder wenn ein Dokument in OneDrive geändert wird. Sogenannte Connectors werden verwendet, um eine Folge von Schritten zu bilden, um den gewünschten Effekt zu erzielen.

Instant Flow

Ein Instant Flow wird mit dem Klicken eines Buttons gestartet. Dieser Flow läuft sowohl auf Mobile als auch auf Desktop Devices. Er wird verwendet, um ein breites Spektrum von Aufgaben zu erledigen, wie die Beantragung einer Genehmigung via Teams oder SharePoint.

Scheduled Flow

Dieser Flow wird nach einem Zeitplan automatisiert und zu einem gewünschten Datum oder beliebiger Uhrzeit ausgeführt werden. Dieser Ablauf ist hilfreich, um zum Beispiel jeden Tag zur selben Uhrzeit einen Daten-Upload zu machen.

Cloud Flow in der Praxis

Die Diplomanden dieser Arbeit entschieden sich für einen Automated Flow, da dieser die gewünschte Funktion erfüllt. Wie bereits erwähnt, ist der Automated Flow ein Event-basierter Flow und wird durch das Erhalten einer E-Mail ausgelöst (Siehe Abbildung 16).

Um alle Anlagen der E-Mail zu verarbeiten, wird eine Schleife um die nächsten Schritte gesetzt. Da es erlaubt ist mehrere Eingangsrechnungen in der E-Mail anzuhängen muss jeder Anhang einzeln verarbeitet werden. Das zuvor definierte AI-Builder Modell extrahiert die notwendigen Information aus dem Dokument (Siehe Abbildung 17).

Bei Eingang einer neuen E-Mail (V3)

Ordner:

An:

Cc:

"An" oder "Cc":

Von:

Anlagen einschließen:

Betrefffilter:

Wichtigkeit:

Nur mit Anlagen:

[Erweiterte Optionen ausblenden](#)

Abbildung 16: Power Automate Flow Trigger

Auf alle anwenden

* Ausgabe von vorherigen Schritten auswählen

Extract information from forms

This action extracts the data from the document you provide when running the flow by using your AI Builder Form Processing model.

* KI-Modell:

* Formulartyp:

* Formular:

[Erweiterte Optionen anzeigen](#)

Abbildung 17: Verarbeitung des Dokuments

Um die extrahierten Daten zu persistieren, wurde eine Dynamics CRM Testumgebung aufgesetzt, um einen echten Geschäftsfall zu simulieren. Es wurden zwei Entitäten, Invoice und Invoice Items erstellt (18). Da die Entität Invoice Items von der Entität Invoice abhängig ist, muss die Entität Invoice Items eigens gespeichert werden.

Kurzer Exkurs; Microsoft Dynamics CRM: Customer Relationship Management (CRM) beinhaltet eine Reihe von Software-Tools, die für das Verwalten der drei Pfeiler zwischen Kunde und Unternehmen zuständig sind: Marketing, Verkauf und Dienstleistungen. Neben Sammeln von Kundendaten aus verschiedenen Quellen und Automatisieren von sich wiederholenden Vertriebs-, Marketing- und Kundendienstprozessen, fördert eine CRM-Software auch die Abteilungsübergreifende Zusammenarbeit. Diese CRM-Software lässt sich in zwei Arten unterscheiden:

1. On-Premise CRM-Software
2. Cloud-basierte CRM-Software

On-Premise CRM-Software

Da das Hauptgeschäft von Unternehmen wie Gesundheits- und Finanzeinrichtungen, der Umgang mit sensiblen Kundendaten ist, wird aus Sicherheitsgründen eine On-Premise Lösung, sprich eine CRM-Software vor Ort, bevorzugt. Diese Systeme sind aber mit hohen Wartungs- und Entwicklungskosten verbunden, und die Unternehmen sind selbst für Sicherheit und Datenpflege zuständig.

Cloud-basierte CRM-Software

Unternehmen haben die Möglichkeit die CRM-Software und die damit verbundenen Aufwände, in die Cloud zu verlagern. Im Gegensatz zu der On-Premise Lösung ist die Cloud-Variante flexibler in Hinsicht auf Speicher- und Rechenressourcen.

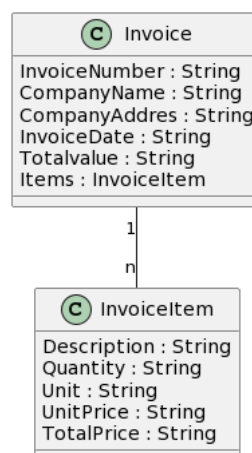


Abbildung 18: Klassendiagramm Invoice

Um die Richtigkeit des Datenmodells zu sichern muss die Reihenfolge der Persistierung beachtet werden und alle Invoice Items abhängig von der dazugehörigen Invoice gemacht werden 19.

The screenshot shows a two-step process for creating a new data record. The top bar indicates 'Neuen Datensatz erstellen (veraltet)' (Create new record (deprecated)). An arrow points down to a second, more detailed form titled 'Neuen Datensatz erstellen (veraltet) 2'. This form is for creating 'Invoice Items' and includes the following fields:

- Organisationsname:** CRM782075
- Entitätsname:** Invoice Items
- Description:** Items Descrip...
- Invoice:** Invoice
- Name:** Bezeichnung
- Quantity:** Items Qty value
- Total Price:** Items Total Pri...
- Unit:** Items Unit val...
- Unit Price:** Items Unit Pri...
- Invoice:** (empty field)

At the bottom, there is a link 'Erweiterte Optionen anzeigen' (Show advanced options) with a dropdown arrow.

Abbildung 19: Persistierung der Entitäten

2.2.2

2.4 Praxis: Cognitive Services

Die Azure Cognitive Services sind ein Teil der Cloud-basierten Dienste von Microsoft. Mithilfe von REST-APIs und Client-SDKs, ist es möglich kognitive Intelligenz in ihre Applikation einzubauen. Ein großer Vorteil ist, dass man dafür wenig bis keine Erfahrung im Bereich der künstlichen Intelligenz und Data Science benötigt. Die, in den Azure Cognitive Services, beinhaltete Sammlung an kognitive Funktionen ermöglicht

es Lösungen zu erstellen, die menschliche Fähigkeiten wie sehen, sprechen und hören nachahmen.

Kurzer Exkurs; REST-API, SDK:

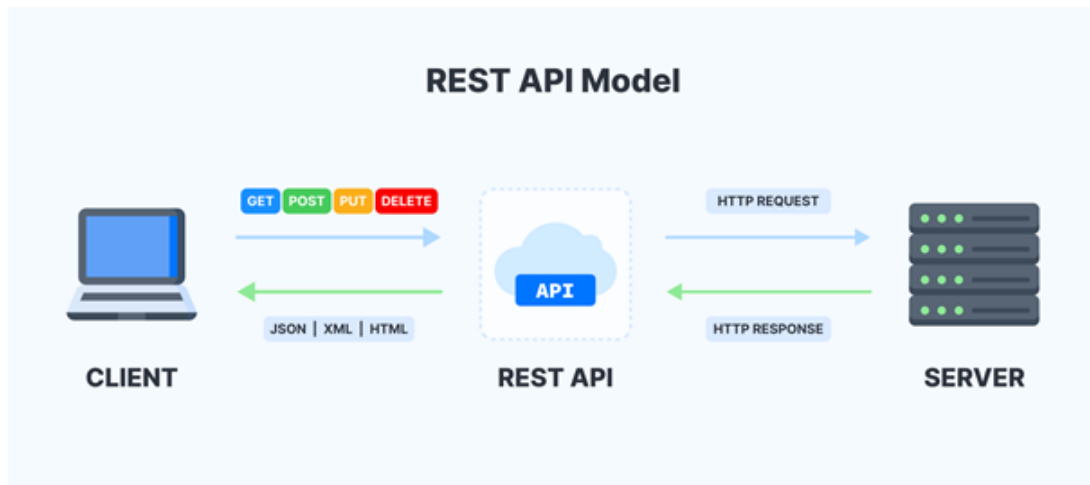


Abbildung 20: Funktionsweise von REST

1. Stabile und interaktive Anwendungen setzen voraus, dass Programme untereinander barrierefrei kommunizieren können. Eine API, Application Programming Interface, definiert regeln, die diese Kommunikation erleichtert. Unter diesen Programmen kann man entweder Softwarebibliotheken, ein Betriebssystem oder einen Webserver verstehen. Der entscheidende Vorteil von APIs sind, dass das anfragende Programm keinerlei Informationen über das dahinterstehende System des Antwortgebers haben muss.

REST-API: Einer der bekanntesten Architekturen einer API ist REST. Representational State Transfer ist ein Stil für die Entwicklung Anwendungen, die untereinander auf irgendeine Weise vernetzt sind. Es wird oft für die Entwicklung von Web-APIs verwendet. Wichtig zu wissen ist, dass REST zustandslos ist. In diesem Kontext bedeutet zustandslos, dass Anfragen des Clients alle notwendigen Informationen für den Server besitzen muss. Zum Beispiel kann der Client nicht davon ausgehen, dass sich der Server an die vorherigen Requests des Clients erinnern kann. REST-APIs senden im Normalfall HTTP-Anfragen wie GET, POST, PUT, etc.

2. Ein **Software Development Kit**, kurz SDK, ist eine Sammlung aus mehreren Tools, die von einem Hersteller einer Programmiersprache, eines Betriebssystems oder einer Hardwareplattform zur Verfügung gestellt werden. Diese Tools können Debugger, Frameworks oder eine Sammlung aus Codebibliotheken sein.

2.4.1 Cognitive Computing

Cognitive Computing ist ein intelligentes System, das durch umfangreiches Lernen und zielgerichtetes Denken mit Menschen in ihrer natürlichen Form spricht und diese nachahmt. Cognitive Computing ist die dritte Ära der Informatik, und gleichzeitig hat Cognitive Computing sowohl in der Wissenschaft als auch in der Industrie breite Aufmerksamkeit auf sich gezogen. Die Kombination aus maschineller und menschlicher Intelligenz kann die komplexesten Probleme der Welt lösen. Die Verarbeitung natürlicher Sprache mit Sentimentanalyse, künstlicher Intelligenz (KI), maschinellem Lernen und neuronalen Netzen ist der Eckpfeiler von Cognitive-Computing-Prozessen, die Probleme wie Menschen lösen können. Heutzutage wenden fortschrittliche Technologien Cognitive Computing in vielen Bereichen an. Angesichts der heutigen Datenexplosion und der sich schnell ändernden Geschäftsumgebungen können kognitive Systeme intelligente, fließende und verbesserte Mensch-Maschine-Interaktionen effektiv angehen. Künstliche Intelligenz wird in vielen Anwendungen wie Alexa, dem Sprachassistenten von Amazon, Netflix und den Algorithmen von Amazon verwendet, um den nächsten Film oder Kauf vorzuschlagen. Einige Beispiele für persönliche Assistenten, die Cognitive Computing verwenden, sind Alexa, Siri, Google Assistant und Cortana. Die Weiterentwicklung dieser Technologie und ihre Übernahme im öffentlichen und privaten Sektor wird die Zukunft des Cognitive Computing aufgrund technologischer Entwicklungspfade und -trends stark beeinflussen. Kognitive Systeme müssen in Geschäfts- und breiten Anwendungen anpassungsfähig, interaktiv, iterativ, zustandsbehaftet und kontextbezogen sein. Zu den Anwendungen, die von solchen Technologien mit Cognitive Computing profitieren können, gehören Finanz- und Investmentfirmen, Gesundheitswesen und Veterinärmedizin, Reisen und Tourismus, Gesundheit und Wellness, Bildung und Lernen, Landwirtschaft, Kommunikations- und Netzwerktechnologie.

2.4.2 Cognitive Computing und Cloud Computing

Cloud Computing virtualisiert die Datenverarbeitung, den Speicher und die Bandbreite. Dadurch werden die Kosten für die Bereitstellung von Softwarediensten gesenkt und die Industrialisierung sowie die Förderung der Anwendung des Cognitive Computings unterstützt. Darüber hinaus bietet die hohe Rechen- und Speicherkapazität des Cloud Computing dynamische, flexible, virtuelle, gemeinsam genutzte und effiziente Rechenressourcendienste für das Cognitive Computing.

Nachdem die Big-Data-Analyse auf der CC-Plattform durchgeführt wurde, werden Technologien wie maschinelles Lernen eingesetzt, um Daten zu analysieren und die Ergebnisse in verschiedenen Bereichen anzuwenden. Die verschiedenen Kategorien von Informationen entsprechen unterschiedlichen Verarbeitungstechnologien. So entsprechen beispielsweise die wörtlichen Informationen der natürlichen Sprachverarbeitung und die bildlichen Informationen dem maschinellen Sehen. Der kognitive Dienst von IBM für Sprache und die kognitive Computing-Anwendung von Google legen den Schwerpunkt auf die Verwirklichung von gehirnähnlicher Wahrnehmung und Urteilsfähigkeit durch den Einsatz eines Cloud-Service-Modells, um präzise Entscheidungshilfen zu bieten. Cloud Computing und das Internet of Things (IoT) bieten eine Software- und Hardwarebasis für Cognitive Computing, während die Big-Data-Analyse Methoden und Denkweisen zur Entdeckung und Erkennung neuer Möglichkeiten und neuer Werte in Daten bereitstellt.

2.4.3 Microsoft Azure Formularerkennung

In der Microsoft Produktfamilie der Azure Cognitive Services werden zahlreiche KI unterstützte Dienste, die zur Verbesserung und Automatisierung von Unternehmenssoftware dient, bereitgestellt. Folgende Dienste für Text- und Bildanalyse stehen zur Verfügung:

1. Form Recognizer
2. Text Analytics
3. Computer Vision
4. Custom Vision

Die Datenextraktion der firmeninternen Rechnungen erfolgt in dieser Arbeit mittels Form Recognizer bzw. Formularerkennung. Der von Microsoft bereitgestellte Form Recognizer ist speziell auf das Extrahieren von einfachen Daten, wie Rechnungsdaten, Rechnungsnummern oder Anschriften der beteiligten Firmen, aber auch komplexere Dokumentstrukturen, wie Tabellen können entnommen werden.

Alle anderen erwähnten Dienste der Cognitive Services, wie Computer Vision, verfügen über die gleiche Grundfunktionalität, unterscheiden sich jedoch in Qualität, Spezifikation und im Ergebnis. Zum Beispiel ist der Dienst Custom Vision in der Lage zur reinen Textextraktion mithilfe von OCR, Bilderkennung und Bewegungsanalysen und der Form

Recognizer ist darauf spezialisiert nur gewünschte, meist firmeninternen, Daten aus einem Formular zu entnehmen.

Für die Nutzung des Form Recognizer Dienstes, wird folgendes vorausgesetzt:

1. Ressourcengruppe
2. Formularerkennung
3. Azure Speicherkonto

Ressourcengruppe

Um die Formularerkennung nutzen zu können, ist die Erstellung einer Ressourcengruppen notwendig. Diese ist zur eindeutigen Zuordnung des zu verwendeten Cloud-Dienstes. Erstellt wird diese im Microsoft Azure Portal unter dem Menüpunkt "Ressourcengruppe erstellen".

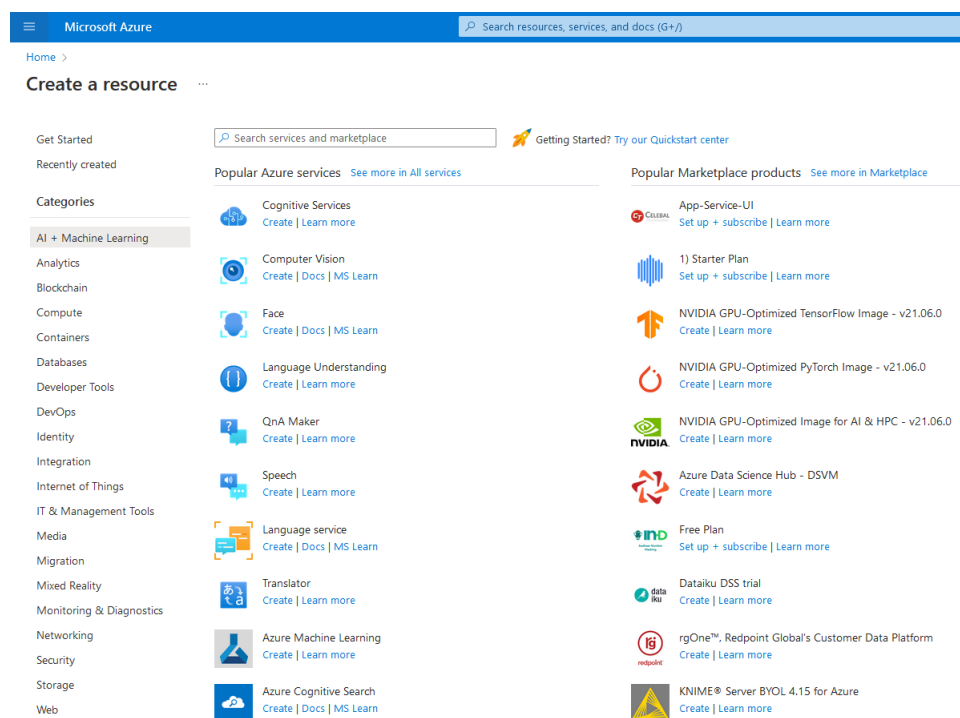


Abbildung 21: Ressourcengruppe erstellen

In der folgenden Abbildung ist zu sehen, dass die Ressourcengruppe "CognitiveServices" genannt wurde und die Serverregion auf "(Europe) Germany West Central" gesetzt wurde. Die Wahl der Serverregion ist insofern relevant, da je nach Region die Verarbeitung der Daten unterschiedlich ist und somit auch die Datenschutzbestimmungen variieren.

The screenshot shows the 'Create a resource group' page in the Microsoft Azure portal. The page is divided into three tabs: 'Basics', 'Tags', and 'Review + create'. The 'Basics' tab is active. Below the tabs, there is a description of a resource group. Under 'Project details', the 'Subscription' is set to 'Azure-Abonnement 1' and the 'Resource group' is set to 'CognitiveServices'. Under 'Resource details', the 'Region' is set to '(Europe) Germany West Central'.

Abbildung 22: Ressourcengruppe Details

2.4.4 Formularerkennung

Im nächsten Schritt wird der eigentliche Cloud-Dienst erstellt. Gleich wie die Ressourcengruppe, ist die Formularerkennungsressource ebenfalls über den Menüpunkt "Ressource erstellen" hinzuzufügen.

Wie in der nachstehenden Abbildung zu sehen ist, wird der Cloud-Dienst der Formularerkennung der zuvor erstellten Ressourcengruppe "CognitiveServices" zugewiesen. Ebenfalls wurde die, ausschließlich für den Umfang des Testzweckes bestehende Preisebene "Standard S0" gewählt.

Azure Speicherkonto

Um das Formularerkennungs-Modell später trainieren zu können, werden Trainingsdaten, bzw. Rechnungen im PDF-Format benötigt. Diese werden in einem dafür erstellten Azure Blob-Speicherkonto festgehalten.

Kurzer Exkurs; Blob bzw. **Binary large object** ist eine große Datei, meist Bild- oder Audiodatei, die wegen ihrer besonderen Größe speziell verarbeitet und gespeichert

Microsoft Azure

Home > Create a resource > Form Recognizer >

Create Form Recognizer

Basics Network Identity Tags Review + create

Accelerate your business processes by automating information extraction. Form Recognizer applies advanced machine learning to accurately extract text, key/value pairs, and tables from documents. With just a few samples, Form Recognizer tailors its understanding to your documents, both on-premises and in the cloud. Turn forms into usable data at a fraction of the time and cost, so you can focus more time acting on the information rather than compiling it.

[Learn more](#)

Project Details

Subscription * ⓘ Azure-Abonnement 1

Resource group * ⓘ CognitiveService
[Create new](#)

Instance Details

Region ⓘ West Europe

Name * ⓘ Diplomarbeit

- ✗ The provided name is already in use. Please pick a different name.
- ✗ The sub-domain name is already used. Please pick a different name.
- ✗ The provided sub-domain name is either invalid or already in use. Please pick a different name.

Information ⓘ The free tier (F0) for this resource type is already being used by your subscription for the resource, DA-Formularerkennung. Click here to view your existing resources or select a paid pricing tier below to proceed.

Pricing tier * ⓘ Standard S0 (1 Call per minute for training API)

[View full pricing details](#)

[Review + create](#) < Previous Next : Network >

Abbildung 23: Ressourcengruppe erstellen

werden muss. Wie der Name bereits verrät werden große Mengen an unstrukturierten binären Daten in einer Blob-Datei gespeichert.

Um ein Speicherkonto zu erstellen müssen die Menüpunkte "Ressource erstellen" und "Speicherkonto" ausgewählt werden. Ebenso wie bei der Formularerkennungsressource, muss die überstehende Ressource "CognitiveServices" verknüpft werden.

Anschließend können alle zuvor erstellten Ressourcen im Menüpunkt "Alle Ressourcen" eingesehen werden, wie in Abbildung 25 zu sehen ist

Im nächsten Schritt wird das Modell erstellt und trainiert um den Dienst Formularerkennung zu nutzen. Zunächst wird das UI zur Erstellung des Modells benötigt. Mit dem nachfolgenden Link ist diese zu erreichen:

<https://fott-2-1.azurewebsites.net/>

Microsoft Azure

Home > Storage accounts >

Create a storage account

Basics | Advanced | Networking | Data protection | Encryption | Tags | Review

[Azure storage accounts](#)

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription * Azure-Abonnement 1

Resource group * CognitiveService [Create new](#)

Instance details

If you need to create a legacy storage account type, please click [here](#).

Storage account name ⓘ * traininginvoices

Region ⓘ * (Europe) Germany West Central

Performance ⓘ *
☒ Standard: Recommended for most scenarios (general-purpose v2 account)
☐ Premium: Recommended for scenarios that require low latency.

Redundancy ⓘ * Geo-redundant storage (GRS)

☒ Make read access to data available in the event of regional unavailability.

Review < Previous Next : Advanced >

Abbildung 24: Ressourcengruppe erstellen

Um ein eigenes Modell zu erstellen, muss, wie in der folgenden Abbildung zu sehen ist, der dritte Punkt "Use Custom to train a model with labels and get key value pairs" ausgewählt werden

Als nächstes muss eine Verbindung zwischen dem zuvor erstellten Azure Blob Speicher-konto und dem Projekt erzeugt werden. Erforderliche Eingaben sind, Verbindungsname und SAS (Shared Access Signature) URI. Diese SAS URI muss davor im Azure Portal in der Übersicht des Azure Blob Speicher-konto generiert werden.

Im nächsten Schritt muss ein sogenannter Security token erstellt werden. Dieser kann im Azure Portal in der Formrecognizerressource erstellt werden. Als Name für den Security token wurde "DA-Formrecognizer Token" gewählt. Siehe nachstehende Abbildung 29.

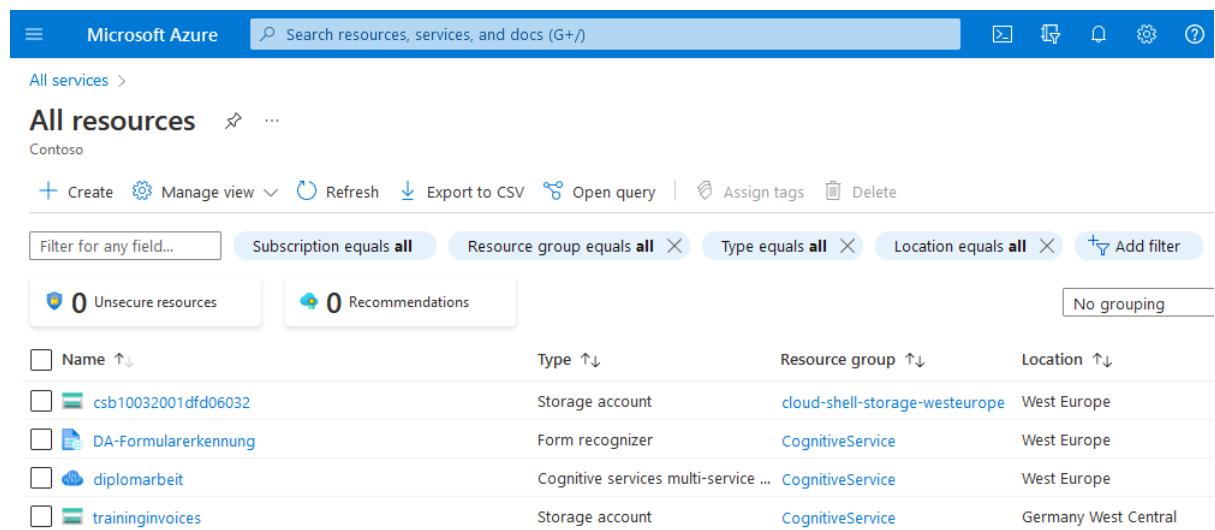


Abbildung 25: Ressourcen im Überblick

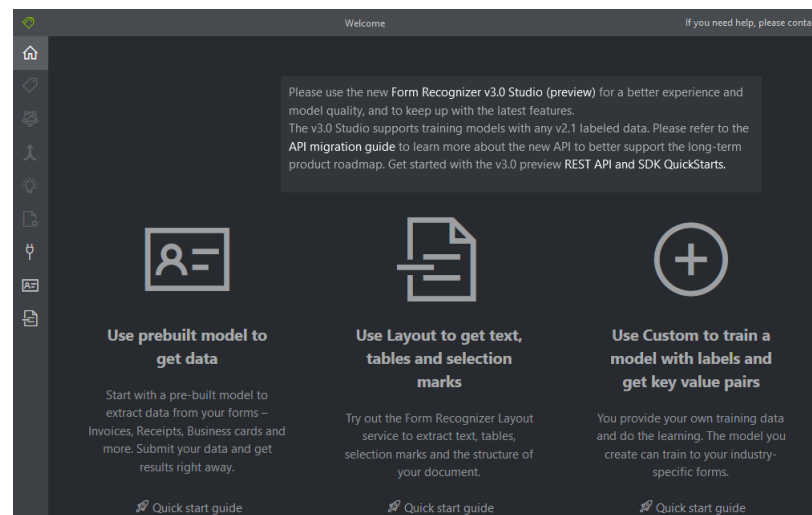


Abbildung 26: Startseite formrecognizer UI

Als finalen Schritt muss jetzt das eigentliche Projekt erstellt werden. Dazu muss im Hauptmenü des Formrecognizer-UI der Punkt "Neues Projekt" gewählt werden. Um ein Projekt zu erstellen sind folgende Eingaben essentiell:

1. Security token
2. Source connection
3. Formrecognizer URL
4. API key

Den zuvor erstellten Security token und connection im Drop-down Menü auswählen und Formrecognizer URL und API key eingeben. Die beiden Eingaben Formrecognizer URL und API key können im Azure Portal in der Formrecognizerressource zu finden.

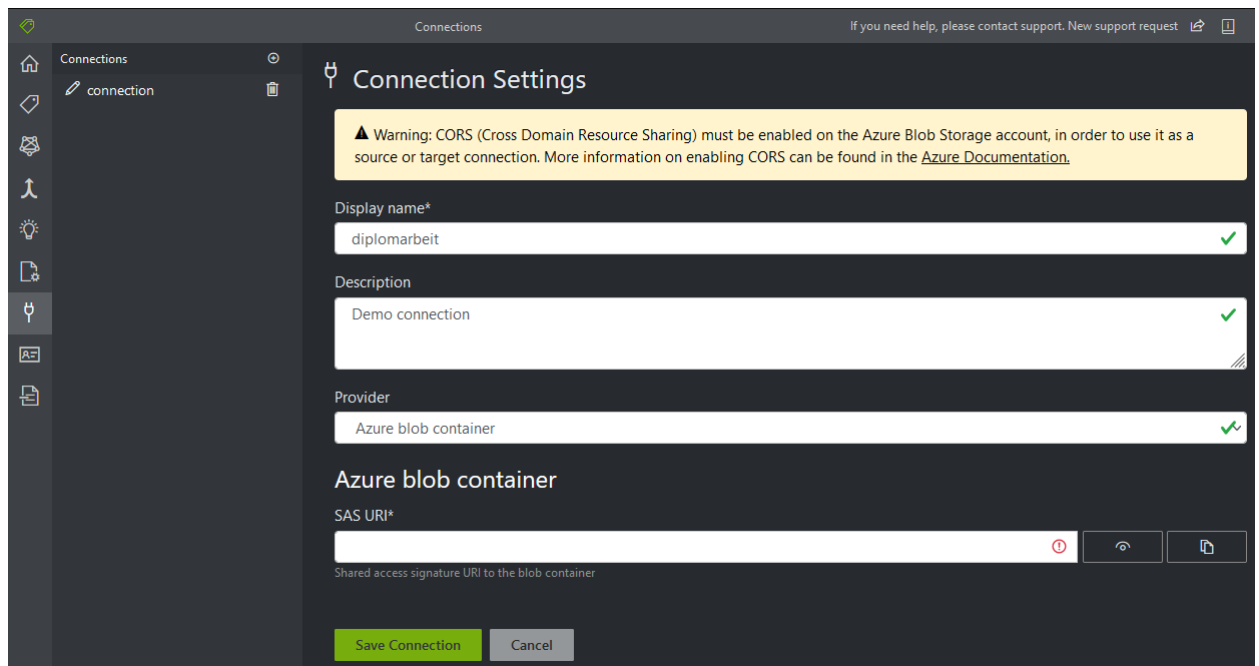


Abbildung 27: Erstellen der Verbindung

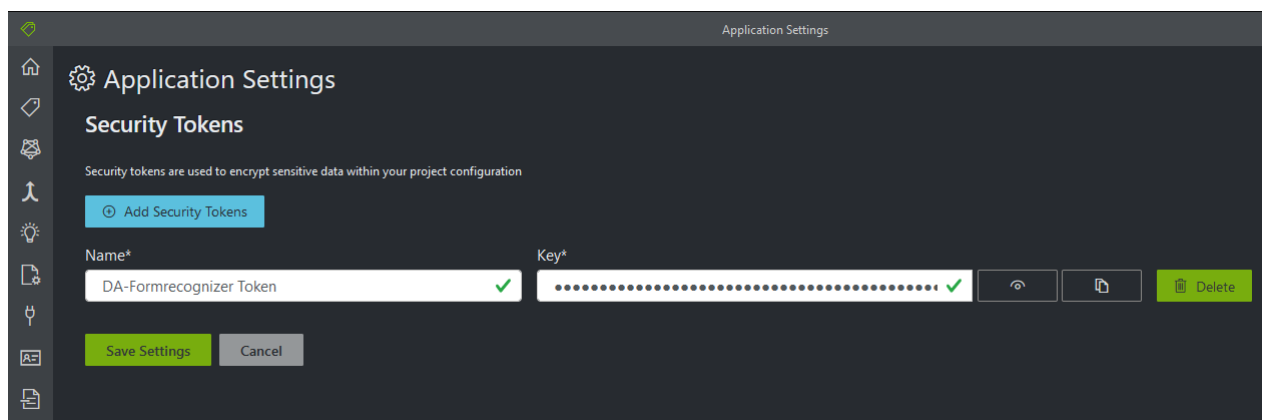


Abbildung 28: Erstellen der Verbindung

Somit ist Erstellung und Konfiguration der Formrecognizerressource abgeschlossen und es kann mit dem Trainieren des Modells begonnen werden.

Trainieren des Modells

Mit den vorherigen Schritten wurde die Konfiguration der Formularerkennungsressource abgeschlossen und in den folgenden Schritten wird das Anlernen an firmeninterne Rechnungen und die darausfolgende Analyse beschrieben. Durch die Auzure Formulaerkennung werden dem Benutzer zwei Möglichkeiten zum Trainieren des Modells zur Verfügung gestellt:

1. das Benutzen von vordefinierten Dokumenten
2. benutzerdefinierte Dokumente

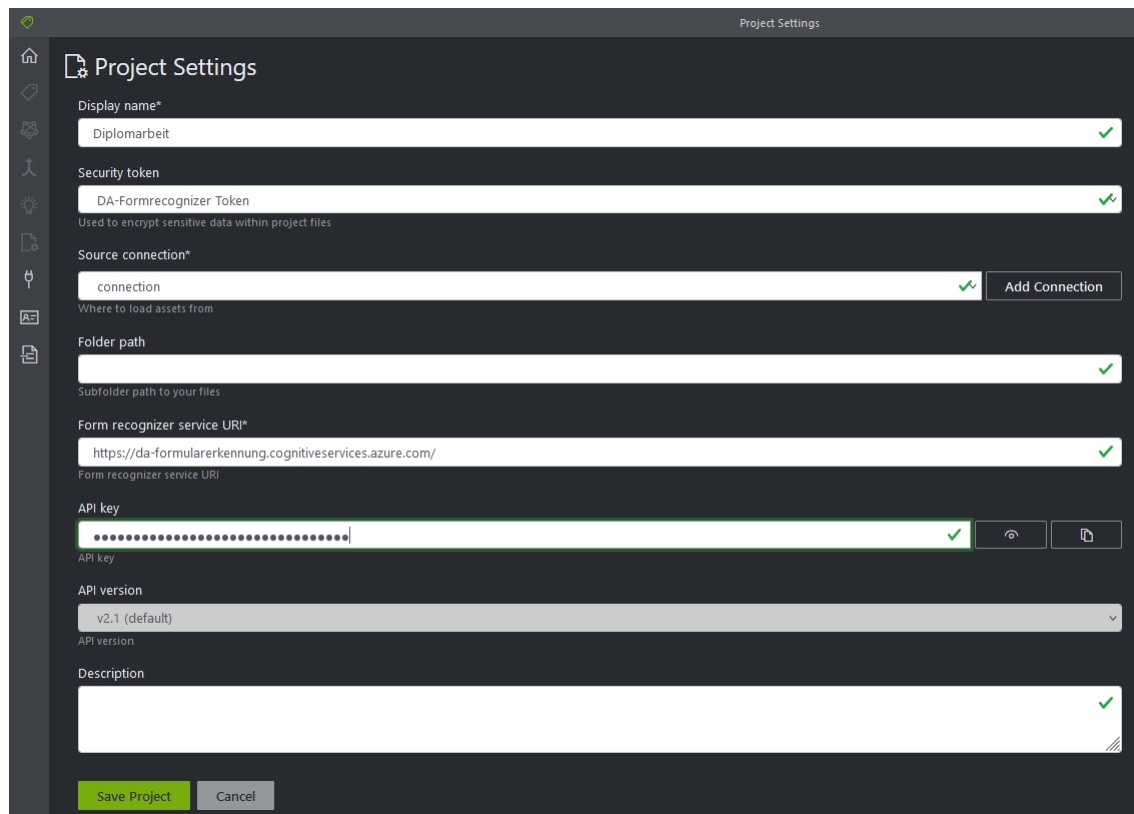


Abbildung 29: Erstellen der Verbindung

Entsprechend der im vorhinein festgelegten Rahmenbedingungen dieser Arbeit, wurden benutzerdefinierte Dokumente, bzw. Rechnungen benutzt. Bei benutzerdefinierten Dokumenten ist der Benutzer im Stande, das Modell auf den Aufbau und Inhalt firmenspezifischer Rechnungen zu trainieren. Das Modell erlernt hierbei, wie Felder bzw. Rechnungselemente zusammenhängen und versucht Muster in den verschiedenen Trainingsdaten zu erkennen und später bei der Analyse anzuwenden.

Es wird von Microsoft empfohlen, für jede Dokumentensammlung fünf oder mehr Dokumente hochzuladen. Um dieser Empfehlung nachzukommen wurden zehn Rechnungssammlungen von Eingangsrechnungen, mit jeweils vier bis sechs Rechnungen zum Trainieren und weitere zwei bis vier zum Testen in den Azure Blob Speicher geladen. Um das Modell trainieren zu können, wurde in den vorherigen Schritten ein Azure Blob Speicherkonto erstellt (2.4.4). Die in dem Blob-Speicher enthaltenen Rechnungen werden automatisch in die Cloud geladen und in dem Formrecognizer UI angezeigt. Die OCR-Software erkennt nach kurzer Wartezeit alle Zeichen und kennzeichnet diese. Weiters müssen anschließend, die von der OCR-Software erkannten Zeichen markiert und den gewünschten Feldern zugewiesen werden. Diese Zuweisung ist essentiell, um dem Modell eine gewisse Dokumentstruktur zu erlernen. Deshalb wurden das Austel-

lungsdatum dem Tag "Invoice Date" und der Gesamtbetrag der Rechnung "TotalValue" zugewiesen. Siehe Abbildung 30

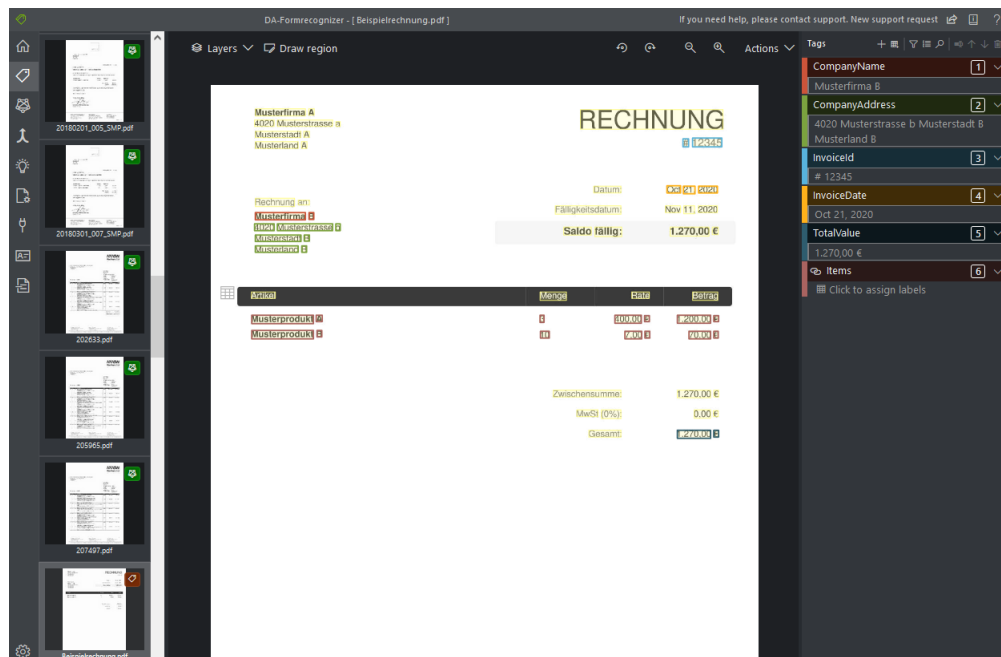


Abbildung 30: Erstellen der Verbindung

Um tabellarische Daten markieren zu können, wird ein spezieller Tag-Typ benötigt, der sogenannte "table tag". Wie in der folgenden Abbildung 32 zu sehen ist, können mehrere Zeilen abermals mit verschiedenen eigenen Tags versehen werden.

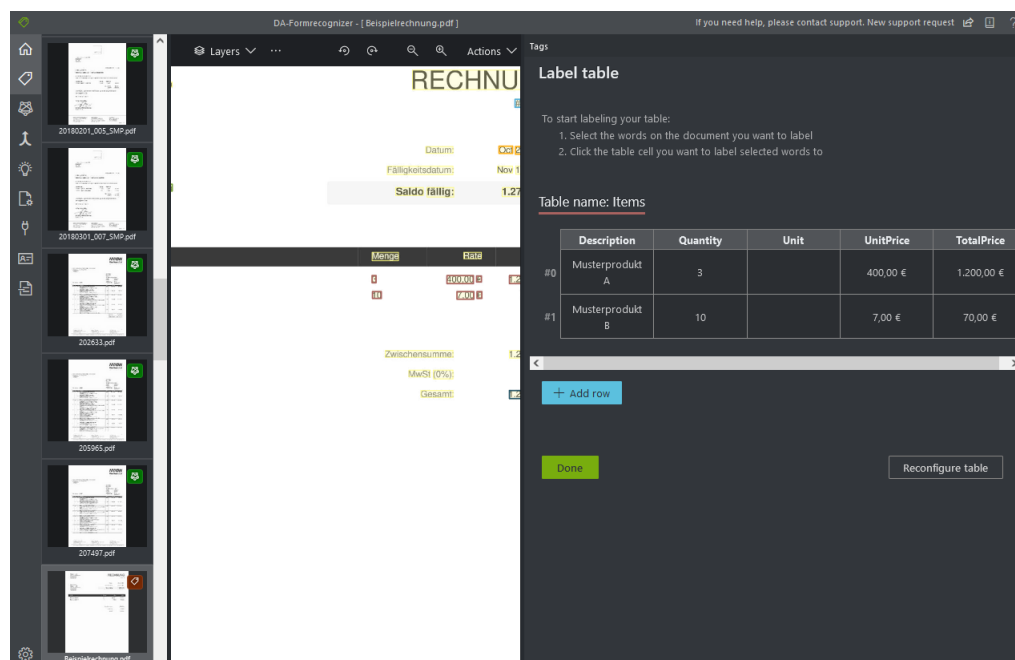


Abbildung 31: Erstellen der Verbindung

Wenn alle Dokumente erfolgreich markiert wurden, kann das Modell im Menüpunkt "Train" trainiert werden. Somit eignet sich die Formularerkennung, anhand der ver-

fügbaren Daten, die Dokumentstruktur jeder einzelnen Rechnungskollektion an. Das Ergebnis wird in der folgenden Abbildung dargestellt.

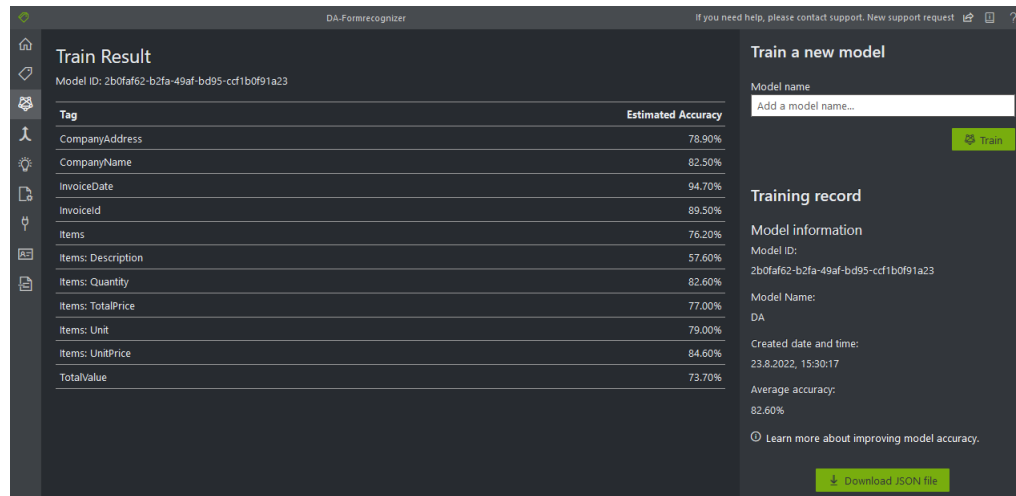


Abbildung 32: Erstellen der Verbindung

Zusammenfassung und Ergebnis der Formulareerkennung

3 Ausgangslage

3.1 Ausgangssituation

3.2 Istzustand

3.3 Problemstellung

3.4 Ziele

3.5 Aufgabenstellung

3.5.1 Funktionale Anforderungen

3.5.2 Nicht funktionale Anforderungen

3.6 Systemarchitektur

3.7 Ablauf

4 Umsetzung

4.1 Konfigurations- und Entwicklungszeit

4.2 Performance und Präzision

4.3 Benutzerfreundlichkeit

4.4 Kosten

5 Gegenüberstellung und Conclusio

5.1 Konfigurations- und Entwicklungszeit

5.2 Performance und Präzision

5.3 Benutzerfreundlichkeit

5.4 Kosten

Literaturverzeichnis

- [1] J. Delua, „Supervised vs. Unsupervised Learning: What’s the Difference? | IBM,” 2021, letzter Zugriff am 03.27.2022. Online verfügbar: <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>
- [2] L. Eikvil, „OCR - Optical Character Recognition,” 1993, letzter Zugriff am 03.27.2022. Online verfügbar: <http://home.nr.no/~eikvil/OCR.pdf>
- [3] G. Shperber, „A gentle introduction to OCR.” 2018, letzter Zugriff am 03.27.2022. Online verfügbar: <https://towardsdatascience.com/a-gentle-introduction-to-ocr-ee1469a201aa>
- [4] E. O. u. Y. W. Boris Epshtein, *Detecting Text in Natural Scenes with Stroke Width Transform*, 2010, letzter Zugriff am 01.04.2022. Online verfügbar: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/201020CVPR20TextDetection.pdf>
- [5] „ICDAR,” letzter Zugriff am 03.30.2022. Online verfügbar: <https://icdar2021.org>

Abbildungsverzeichnis

1	Evolution von Künstlicher Intelligenz	1
2	Feature Extraction, ML vs DL	2
3	Vernetzungen nach der Geburt, nach 3 Monaten und nach 15 Monaten	4
4	Decision Tree an dem Beispiel von 2	7
5	Unkategorisierte Daten	8
6	Kategorisierte Daten	8
7	Bouning Boxes Beispiel	11
8	Die Kanten des Striches (a) werden solange verglichen, bis zwei gefunden werden mit der mit der gleichen Richtung (b). Alle unterliegenden Pixel erhalten die Strichbreite der Entfernung zwischen der Start- und Endkante (c).	12
9	CC-Overview	14
10	Modelle	16
11	Anwendungsfälle	17
12	AI-Builder Modelle	20
13	Beispiel Eingansrechnung	22
14	AI-Builder Felder	23
15	AI-Builder Felder Tagging	24
16	Power Automate Flow Trigger	26
17	Verarbeitung des Dokuments	26
18	Klassendiagram Invoice	27
19	Persistierung der Entitäten	28
20	Funktionsweise von REST	29
21	Ressourcengruppe erstellen	32
22	Ressourcengruppe Details	33
23	Ressourcengruppe erstellen	34
24	Ressourcengruppe erstellen	35
25	Ressourcen im Überblick	36
26	Startseite formrecognizer UI	36
27	Erstellen der Verbindung	37
28	Erstellen der Verbindung	37
29	Erstellen der Verbindung	38
30	Erstellen der Verbindung	39
31	Erstellen der Verbindung	39
32	Erstellen der Verbindung	40

Tabellenverzeichnis

1	Beispiel einer Zwicky-Boy	4
2	Beispiel einer Ausprägung	5
3	Ergebnis einer Ausprägung	5
4	Ähnlichkeiten zwischen Buchstaben im Lateinischen und Kyrillischen Alphabet	10

Quellcodeverzeichnis

Anhang