

# Bachelorarbeit

Yavuz Arslan  
Evaluierung cloudbasierter Machine Learning  
Services

Yavuz Arslan

# Evaluierung cloudbasierter Machine Learning Services

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung  
im Studiengang Bachelor of Science Angewandte Informatik  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Ulrike Steffens  
Zweitgutachter: Prof. Dr. Stefan Sarstedt

Eingereicht am: 11. Januar 2019

**Yavuz Arslan**

**Thema der Arbeit**

Evaluierung cloudbasierter Machine Learning Services

**Stichworte**

MLaaS, Maschinelles Lernen, Cloud, Gesichtserkennung, Sentimentanalyse

**Kurzzusammenfassung**

Die Popularität maschinellen Lernens steigt rasant. Immer mehr Unternehmen nutzen dessen Potenzial, um aus Daten wertvolle Vorhersagen zu generieren. Cloudbasierte Dienste bieten dabei die Chance, die Vorteile maschinellen Lernens leicht und schnell in eigene Anwendungen zu integrieren und gleichzeitig die Vorzüge des Cloud Computing zu nutzen. Diese Thesis untersucht die cloudbasierten Machine Learning Services von Amazon, Google, IBM und Microsoft. Für die Dienste der Gesichts- und Sentimentanalyse werden außerdem Tests konzipiert, um die Anbieter untereinander zu vergleichen und sie anhand ihrer Qualität zu bewerten.

**Title of Thesis**

Evaluation of cloud-based machine learning services

**Keywords**

MLaaS, Machine Learning, Cloud, Face Detection, Sentiment Analysis

**Abstract**

Popularity in machine learning rises rapidly as more and more companies use its capabilities to generate useful predictions from data. Cloud-based services offer the opportunity to leverage the benefits of machine learning by providing a simple and quick way to integrate them into applications and simultaneously profit by the advantages of cloud computing. This thesis explores the cloud-based machine learning services from Amazon, Google, IBM and Microsoft. Subsequently, the respective face detection and sentiment analysis services are compared against each other to evaluate their qualities.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>v</b>
<b>Tabellenverzeichnis</b>	<b>vi</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Ziel . . . . .	1
1.2 Abgrenzung . . . . .	1
1.3 Struktur . . . . .	2
<b>2 Grundlagen</b>	<b>3</b>
2.1 Cloud Computing . . . . .	3
2.2 Maschinelles Lernen . . . . .	5
2.2.1 ML Modell . . . . .	5
2.2.2 Supervised Learning . . . . .	5
2.2.3 Unsupervised Learning . . . . .	6
2.2.4 Künstliche Neuronale Netze . . . . .	7
2.3 Cloudbasierte ML Services . . . . .	8
2.3.1 REST API . . . . .	8
2.3.2 Jupyter Notebook . . . . .	8
2.3.3 SDK . . . . .	9
2.3.4 API Key . . . . .	9
2.3.5 JSON . . . . .	9
2.3.6 Frameworks . . . . .	9
<b>3 Angebotsübersicht</b>	<b>10</b>
3.1 Plattformbasierte Dienste . . . . .	10
3.2 Vortrainierte ML APIs . . . . .	12
3.2.1 Sprach- & Textanalyse . . . . .	12
3.2.2 Bildanalyse . . . . .	16

3.2.3	Sonstige Services . . . . .	18
3.3	(Semi-)Automatisierte Dienste . . . . .	18
<b>4</b>	<b>Evaluierung</b>	<b>20</b>
4.1	Testumgebung . . . . .	20
4.2	Bildanalyse . . . . .	21
4.2.1	Datensatz . . . . .	21
4.2.2	Aufbau . . . . .	22
4.2.3	Bewertungskriterien . . . . .	25
4.2.4	Auswertung . . . . .	27
4.3	Textanalyse . . . . .	28
4.3.1	Datensatz . . . . .	28
4.3.2	Aufbau . . . . .	29
4.3.3	Bewertungskriterien . . . . .	31
4.3.4	Auswertung . . . . .	32
<b>5</b>	<b>Resümee</b>	<b>39</b>
5.1	Zusammenfassung . . . . .	39
5.2	Ausblick . . . . .	41
5.3	Reflexion . . . . .	42
	<b>Selbstständigkeitserklärung</b>	<b>49</b>

# Abbildungsverzeichnis

2.1	Ablauf <i>Supervised Learning</i> . . . . .	6
2.2	Aufbau KNN . . . . .	7
4.1	Beispielbilder aus dem LFW Datensatz . . . . .	21
4.2	Verteilung der Produktbewertungen . . . . .	29
4.3	GCP Sentiment Verteilung . . . . .	33
4.4	Azure Sentiment Verteilung . . . . .	34
4.5	IBM Sentiment Verteilung . . . . .	34
4.6	Sentiment Verteilung für 3er Rezensionen bei GCP . . . . .	35
4.7	Sentiment Verteilung für 3er Rezensionen bei Azure . . . . .	36
4.8	Sentiment Verteilung für 3er Rezensionen bei IBM . . . . .	36
4.9	Sentiment Verteilung bei AWS . . . . .	37

# Tabellenverzeichnis

3.1	Übersicht Plattformbasierte Dienste . . . . .	11
3.2	Funktionsübersicht Sprach- & Textanalyse . . . . .	13
3.3	Übersicht Bilderkennung . . . . .	17
3.4	Übersicht autom. Dienste . . . . .	19
4.1	Wahrheitsmatrix Gesichtserkennung . . . . .	27
4.2	Gesichtserkennung Ergebnisse . . . . .	27
4.3	Geschlechtserkennung Ergebnisse . . . . .	28
4.4	Beispiel Bewertung . . . . .	28
4.5	Sentiment-Kriterien für die Kategorisierung . . . . .	32
4.6	Ergebnisse der Kategorisierung nach positivem oder negativem Sentiment	37
4.7	Laufzeiten Sentimentanalyse . . . . .	38

# Listings

4.1	AWS Annotationfunktion . . . . .	23
4.2	Beispiel Azure Annotierung . . . . .	24
4.3	Beispiel Antwort GCP Sentimentanalyse . . . . .	31



# 1 Einleitung

Heutzutage erheben und verarbeiten immer mehr Unternehmen große Datenmengen. Zahlreiche neue datengetriebene Geschäftsmodelle sind entstanden. Die Nachfrage ist groß, jedoch ist die Verarbeitung der Daten ein mühsamer und zeitintensiver Prozess. Neue Einsichten und Handlungsstrategien anhand der riesigen Datenmengen abzuleiten stellt eine große Hürde da. Maschinelles Lernen soll dabei Abhilfe schaffen. Anwendungen können mit dessen Hilfe Modelle erstellen, die die Daten in kürzester Zeit verarbeiten und wichtige Erkenntnisse generieren. Meist setzt der Einsatz vom maschinellen Lernen jedoch tiefgreifende mathematische Fachkenntnisse voraus; das Aufsetzen eigener Anwendungen ist zeitraubend und kostspielig. Um diese Probleme zu umgehen werden cloudbasierte Machine Learning Services angeboten. Diese Dienste abstrahieren den großen Mehraufwand und bieten zusätzlich die bereits bekannten und stark gefragten Vorteile des Cloud Computing.

## 1.1 Ziel

In dieser Arbeit sollen die gängigsten cloudbasierten Machine Learning Services von vier ausgewählten Anbieter näher untersucht und evaluiert werden. Dabei soll zudem veranschaulicht werden, welche Arten von Diensten verfügbar sind und welche Anwendungsbereiche sie abdecken.

## 1.2 Abgrenzung

Es gibt zahlreiche Unternehmen, die cloudbasierte Machine Learning Dienste anbieten. Da nicht das gesamte Angebot im Rahmen dieser Arbeit untersucht werden kann, beschränkt sich diese auf die Dienste von Amazon, Google, IBM und Microsoft. Die Dienste dieser vier Anbieter werden von Unternehmen am meisten für ihre Anwendungen genutzt

(RightScale, Inc. [2018]). Der Fokus liegt auf den technischen Funktionen der Dienste; wirtschaftliche Aspekte wie Kosten oder Service-Level-Agreements werden nicht untersucht, ebenso wurden Sicherheitsaspekte nicht näher betrachtet.

Eine vollständige Evaluierung über alle Machine Learning relevanten Dienste der Anbieter ist im Umfang dieser Arbeit nicht möglich. Daher konzentriert sich die im weiteren Verlauf folgende Evaluierung auf einen Teilbereich der Dienste. Die Auswahl der jeweiligen Dienste wird zu Beginn der Evaluierung erläutert.

### 1.3 Struktur

Zu Beginn der Arbeit werden in Kapitel 2 die notwendigen Grundlagen vermittelt, die für ein besseres Verständnis der darauffolgenden Abschnitte sorgen. Der Hauptteil der Arbeit ist in zwei Abschnitte unterteilt. In dem ersten Abschnitt (Kapitel 3) wird eine Angebotsanalyse durchgeführt. Dort werden sämtliche ML relevanten Cloud-Dienste der Anbieter aufgeführt, kategorisiert und näher untersucht.

Im darauf folgenden Kapitel 4 wird zunächst erläutert, welche Dienstkategorien sich für eine Evaluierung eignen. Anschließend werden die geeigneten Dienste ausgewählt und gegenüber gestellt. Hierfür werden Tests konzipiert, anhand deren eine Bewertung der Anbieter erfolgt. Abschließend werden in Kapitel 5 die gesammelten Ergebnisse aus den beiden Hauptteilen zusammengefasst und es folgen ein kurzer Ausblick in die Zukunft von cloudbasierten ML Services sowie eine Reflexion der Arbeit.

## 2 Grundlagen

In diesem Kapitel werden die wesentlichen Eigenschaften maschinellen Lernens und cloud-basierter Dienste näher erläutert. Dies gewährt einen groben Überblick über die beiden Komponenten von cloudbasierten Machine Learning Services. Zudem werden relevante Technologien vorgestellt, die bei der Nutzung von cloudbasierten Machine Learning Services zum Einsatz kommen.

### 2.1 Cloud Computing

*Cloud Computing* hat in den vergangenen Jahren eine bedeutende Rolle in der digitalen Transformation eingenommen. Unternehmen setzen auf Cloud Technologien und passen ihre Geschäftsprozesse dementsprechend an. Auch in Zukunft sollen das Interesse und die Einnahmen im Cloud Geschäft wachsen (Gartner, Inc. [2018]).

*Cloud Computing* als Ganzes eindeutig zu definieren fällt schwer, da sich keine Definition als allgemeingültig durchsetzen konnte. Eine der geläufigsten und viel anerkannten Definitionen ist die der US-amerikanischen Standardisierungsstelle NIST (National Institute of Standards and Technology) (Mell and Grance [2011]), welche auch von der ENISA (European Network and Information Security Agency) genutzt wird:

„Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.“

Weiterhin benennt das NIST fünf Eigenschaften, welche einen Cloud-Dienst charakterisieren:

- *On-demand self-service*: Nutzer können selbständig Rechenkapazitäten (z. B. Rechenleistung, Speicher) einrichten und verwalten, ohne dass eine menschliche Interaktion mit dem Anbieter notwendig ist
- *Broad network access*: Dienste sind über das Netz erreichbar und erlauben Zugriff über standardisierte Mechanismen, welche nicht an bestimmte Client-Typen (z.B. PC, Smartphone, etc.) gebunden sind.
- *Resource pooling*: Die Ressourcen des Anbieters liegen in einem Pool vor und können durch ein mandantenfähiges System von mehreren Nutzer dynamisch genutzt werden. Die Ressourcen sind dabei ortsunabhängig, sodass die Nutzer in der Regel über keine Kontrolle oder kein Wissen von dem exakten Standort der Ressourcen verfügen. Jedoch besteht die Möglichkeit, den Standort auf einer höheren Ebene (z.B. Land oder Kontinent) festzulegen.
- *Rapid elasticity*: Dienste können elastisch zu- und abgeschaltet werden (u.a. automatisch), so dass eine schnelle Skalierbarkeit gegeben ist. Für die Nutzer ist meist kein Limit bezüglich der nutzbaren Kapazitäten erkennbar.
- *Measured service*: Cloud Systeme kontrollieren und optimieren automatisch die Ressourcennutzung mit Hilfe von Metriken. Die Ressourcennutzung kann transparent sowohl vom Anbieter als auch vom Nutzer eingesehen und kontrolliert werden.

Neben diesen Eigenschaften werden drei Haupt-Dienstmodelle spezifiziert:

- *Software as a Service (SaaS)*: Unter diese Dienste fallen Anbieter-Applikationen, die auf der Cloud Infrastruktur laufen. Die Applikationen sind entweder über Software Clients (z.B. Webbrowser) oder Programm-Schnittstellen erreichbar. Der Nutzer verwaltetet und kontrolliert nicht die darunter liegende Infrastruktur.
- *Platform as a Service (PaaS)*: Diese Dienste erlauben Nutzern selbst erstellte oder erworbene Applikationen auf der Cloud Infrastruktur zu starten. Der Nutzer hat keine Kontrolle über die genutzte Infrastruktur, kann jedoch die Applikationen und ggfs. die Anwendungsumgebung konfigurieren.
- *Infrastructure as a Service (IaaS)*: Nutzer können eigenständig Rechenkapazitäten, Speicher, Netzwerke und andere Rechenressourcen provisionieren und auf diesen beliebige Software starten (z.B. Applikationen oder Betriebssysteme).

## 2.2 Maschinelles Lernen

*Maschinelles Lernen (Machine Learning, ML)* ist ein Teilbereich der Informatik und eine Anwendungsmöglichkeit der künstlichen Intelligenz. ML fokussiert sich hauptsächlich auf das Entwickeln von Anwendungen/Algorithmen, welche sich eigenständig anhand neuer Daten abwandeln und verbessern.

Täglich werden über 2,5 Exabyte an Daten generiert (Marr [2018]) und viele davon liegen in unstrukturiert Form da (Audio, Videos, Dokumente, etc.). Es ist ein großes Problem diese Daten schnell und effektiv zu verarbeiten. ML bietet die Möglichkeit, dieses Datenvolumen in möglichst kurzer Zeit zu untersuchen, um so Muster zu erkennen und Erkenntnisse zu generieren. Diese Erkenntnisse lassen sich durch ML Anwendungen verallgemeinern und können so für neue Problemlösungen oder für die Analyse von bisher unbekanntem Daten verwendet werden.

### 2.2.1 ML Modell

Ein ML *Modell* ist ein mathematisches Modell, welches anhand von Daten Vorhersagen erstellt. Diese Modelle werden mit Hilfe von Lern-Algorithmen und Trainings- und Testdaten erzeugt. Eine ML Anwendung versucht während der Nutzung dieses Modell mittels neuer Daten dahingehend zu optimieren (weiter zu trainieren), sodass anhand vorgegebener Metriken bessere Vorhersagen erstellt werden.

ML Techniken werden in unterschiedliche Kategorien eingeteilt. Die im *Cloud Computing* verbreitetsten Kategorien sind *Supervised* und *Unsupervised Learning*.

### 2.2.2 Supervised Learning

Bei dem *Supervised Learning (Überwachtes Lernen)* werden Modelle trainiert, die zu einem Eingabewert  $\mathcal{X}$  den dazu vermuteten Ergebniswert  $\mathcal{Y}$  vorhersagen (ähnlich einer mathematischen Abbildung). Dabei sind die möglichen Wertebereiche der Ergebnisse im Voraus bekannt. Die Trainingsdaten bestehen aus Eingaben und den dazu bereits bekannten Ergebnisdaten. Während des Lernprozesses können die Vorhersagen anhand der bekannten Ergebnisse „überwacht“ werden. Die bekannten Ergebnisse einer Eingabemenge werden als *Ground Truth (GT)* bezeichnet.

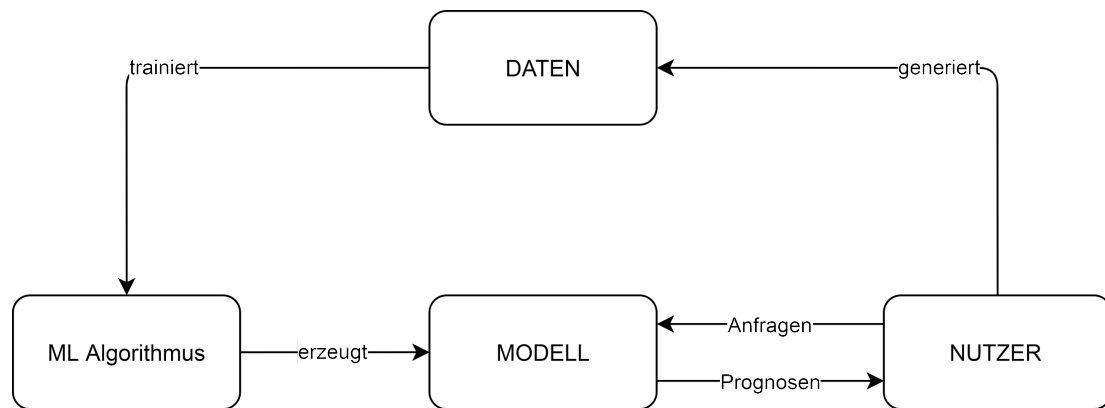


Abbildung 2.1: Ablauf *Supervised Learning*

*Supervised Learning* ist ein zwei Analyseverfahren unterteilt:

### Regression

Die Regression gibt die Beziehung zwischen einer Eingabe- und Ausgabevariable an. Dabei liegen die Ergebniswerte in einer kontinuierlichen Verteilung vor. Ein Beispiel für die Regression ist die Mietpreis-Vorhersage, wo der Preis für eine Wohnung von Eingabewerten wie der Anzahl der Räume, der Lage oder dem Baujahr abhängt.

### Klassifizierung

Bei der Klassifizierung liegen die Ergebniswerte in einer diskreten Form vor, wie z.B. Kategorien. Falls lediglich zwei unterschiedliche Ergebnis-Kategorien vorhanden sind, spricht man von einer Binären-Klassifizierung; bei mehr Kategorien von Mehrklassen-Klassifizierung. Ein bekanntes Beispiel für die (Binäre-)Klassifizierung ist die Erkennung von Spam Mails, bei der ein Klassifizierungsmodell (auch Klassifikator genannt) entscheidet, ob eine Mail als Spam markiert wird oder nicht.

### 2.2.3 Unsupervised Learning

Bei dem *Unsupervised Learning (Unbewachtes Lernen)* sind im Gegensatz zum *Supervised Learning* die Ergebniswerte der Vorhersagen nicht bekannt. Die Trainingsdaten sind nicht klassifiziert/kategorisiert. Der Lern-Algorithmus versucht mit Hilfe der Daten Strukturen

zu erkennen und anhand derer die Daten zu kategorisieren. Einer der verbreitetsten Anwendungsfälle ist *Clustering*. Dabei werden ähnliche Entitäten zusammen gruppiert und so unterschiedliche Cluster erzeugt. Das Modell soll anschließend für neue Daten vorhersagen, in welche Cluster sie passen. Ein Anwendungsfall ist das Gruppieren von Kunden basierend auf ähnlichem Einkaufsverhalten.

### 2.2.4 Künstliche Neuronale Netze

*Künstliche Neuronale Netze* (KNN) werden bereits seit geraumer Zeit im wissenschaftlichen Kontext eingesetzt. Seit ein paar Jahren erfreuen sich KNN auch in der Wirtschaft großer Beliebtheit, nicht zuletzt dank der enormen Fortschritte in der Hardware für Parallele Programmierung und den enormen Datenmengen, die heutzutage verfügbar sind.

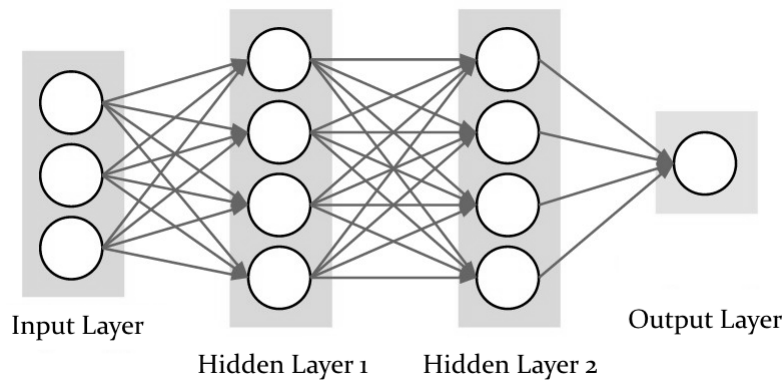


Abbildung 2.2: Aufbau KNN

KNN bestehen aus mehreren einzelnen Informationsverarbeitungseinheiten, den sogenannten **Neuronen**. Die Neuronen verteilen sich auf mehrere Schichten (*Layers*), welche die Tiefe des Netzes angeben, und sind miteinander verbunden. Wie die Verbindungen und Schichten aufgebaut sind hängt von der genutzten Architektur des Netzes ab. Die Informationen fließen von der Eingabeschicht (*Input Layer*) über eine oder mehrere Zwischenschichten (*Hidden Layer*) bis hin zur Ausgabeschicht (*Output Layer*). Dabei ist der Output von einem Neuron der Input des Nächsten. *Deep Learning* Verfahren ermöglichen das effiziente trainieren tiefer Netze in kürzester Zeit. Durch die Masse an Daten können Modelle mit tausenden von Parametern aufgesetzt werden.

## 2.3 Cloubasierte ML Services

*Cloubasierte ML Services (Machine Learning as a Service, MLaaS)* ist ein Sammelbegriff für alle ML relevanten Cloud Computing Dienste. Dabei kommen entweder ein oder mehrere ML basierte Technologien zum Einsatz oder es handelt sich um Dienste, bei denen der Fokus auf der Erstellung von ML Anwendungen liegt. *MLaaS* Dienste können abhängig von ihren Eigenschaften als *SaaS* oder *PaaS* Dienste eingestuft werden.

Für die Nutzung und Anbindung cloubasierter ML Dienste kommen unterschiedliche Technologien zum Einsatz. Im Folgenden werden verbreitete Technologien vorgestellt, die unter anderem für die Evaluierung der Dienste genutzt wurden.

### 2.3.1 REST API

*REST* steht für *Representational State Transfer* und ist ein von Fielding [2000] entworfener Softwarearchitekturstil, welcher Prinzipien definiert, die für die Entwicklung von Webservices genutzt werden. Durch die Anforderung an eine einheitliche Schnittstelle bieten RESTful Webservices systemübergreifende Kompatibilität und eignen sich so hervorragend für Cloud Umgebungen. Diese Schnittstellen werden als REST APIs bzw. RESTful APIs bezeichnet. In den meisten Fällen verwenden diese Schnittstellen HTTP als Protokoll, wodurch keine zusätzlichen Bibliotheken oder Anwendungen zur Nutzung notwendig sind. Datenzugriffe erfolgen über URLs. Jede URL stellt eine Anfrage dar und die zurückgegebenen Daten die Antwort. REST APIs sind die am häufigsten angebotenen Schnittstellen für Cloud-Dienste.

### 2.3.2 Jupyter Notebook

*Jupyter Notebook* ist eine webbasierte, interaktive Entwicklungsumgebung für über 40 Programmiersprachen (Project Jupyter [2019]). Neben dem Programmieren ermöglichen die Notebooks auch Datenanalyse und -visualisierung. Sie vereinfachen durch ihre Handhabung das Prototyping neuer Anwendungen und eignen sich gut für kollaborative Projekte, da die Notebooks mandantenfähig sind und auch leicht über unterschiedliche Exportformate ausgetauscht werden können. Zahlreiche Cloud-Anbieter unterstützen Jupyter Notebook als Nutzerinterface - vor allem im Bereich ML - für ihre Dienste an.



### 2.3.3 SDK

Ein *SDK (Software development kit)* ist eine Sammlung von Software-Entwicklung-Tools zur Erstellung bestimmter Software. Diese können für eine spezielle Hardware oder Plattform zusammengesetzt sein. Einige Cloud Anbieter offerieren SDKs für unterschiedliche Programmiersprachen und bieten so eine alternative Schnittstelle für ihre Dienste an (im Vergleich zu REST APIs).

### 2.3.4 API Key

API Keys werden für den authentifizierten Zugriff auf Webservices genutzt und sind auch beim Cloud Computing weit verbreitet. Der API Key dient zugleich als Identifizierungsmerkmal und Sicherheitstoken für die Authentifizierung und ist in den meisten Fällen einem speziellen Dienst zugeordnet. Falls API Keys verwendet werden, muss bei der Nutzung eines Dienstes der dazugehörige API Key mit übergeben werden.

### 2.3.5 JSON

*JavaScript Object Notation (JSON)* ist ein standardisiertes Datenaustauschformat, welches für Menschen leicht zu lesen ist. Es ermöglicht den Austausch von serialisierbaren Objekten. JSON basiert auf zwei Datenstrukturen: Sammlungen von Key-Value Paaren und geordneten Listen von Werten. Bei den meisten REST APIs wird JSON für die Darstellung von Nutzdaten verwendet.

### 2.3.6 Frameworks

*Frameworks* dienen als Programmiergerüst für Entwickler. Diese sind domänenspezifisch und vereinfachen die Entwicklung von Anwendungen in ihrem jeweiligen Einsatzbereich. Sie beinhalten Bibliotheken mit wohldefinierten Schnittstellen, durch die Teile der Anwendungslogik abstrahiert werden. Heutzutage gibt es zahlreiche Frameworks für ML, mit denen Entwickler schnell und einfach eigene Modelle erstellen und trainieren können. Dadurch ist zum einen weniger Fachwissen im Bereich ML notwendig und zum anderen entfällt ein großer Programmieraufwand.

## 3 Angebotsübersicht

Es gibt zahlreiche Unternehmen, die cloudbasierte Machine Learning Dienste anbieten. Viele dieser Anbieter fokussieren sich dabei auf ein spezielles Einsatzgebiet, wie z.B. Sprachanalyse oder Predictive Analytics.

Diese Arbeit konzentriert sich auf vier führende und bekannte Cloud Plattformen: Amazon Web Services (AWS), Google Cloud Platform (GCP), IBM Cloud/Watson und Microsoft Azure.

Cloudbasierte ML Dienste lassen sich aktuell anhand ihren technischen Anforderungen, der Stack-Tiefe und ihrer Anwendungsgebiete in drei verschiedene Kategorien einordnen. Im Folgenden werden diese Kategorien näher erläutert und die jeweiligen Dienstleistungen der Anbieter analysiert.

### 3.1 Plattformbasierte Dienste

Plattformbasierte Dienste bieten eine verwaltete Plattform (PaaS) für ML, mit der Entwickler möglichst schnell und einfach ihre eigenen Modelle trainieren und einsetzen können. Dabei steht ihnen eine Auswahl unterschiedlicher Frameworks und Algorithmen zur Verfügung.

Durch die Abstraktion der Plattformschicht entfällt der Mehraufwand für die Verwaltung eigener Hardware. Zusätzlich bietet eine cloudbasierte Infrastruktur auch Vorteile in der Skalierbarkeit und Elastizität. Diese Dienste bieten einem die größte Flexibilität hinsichtlich Modellaufbau und Tests anhand unterschiedlicher Algorithmen.

Die wichtigsten Faktoren bei der Auswahl eines solchen Dienstes sind die verfügbaren Frameworks, die unterstützten Programmiersprachen und die Anbindung an den Dienst.

	Amazon SageMaker	Google Cloud ML Engine	Azure ML Services	IBM Watson Studio
Schnittstellen	CLI REST Jupyter Notebook	CLI REST	CLI Jupyter Notebook	CLI Jupyter Notebook
SDKs	.NET, C++, Go, Java, JavaScript, PHP, Python, Ruby	C#, Go, Java, Node.js, PHP, Python, Ruby	Python	Python
Feat. Frameworks	Apache MXNet XGBoost TensorFlow PyTorch Chainer	scikit-learn XGBoost TensorFlow	Apache MXNet scikit-learn TensorFlow PyTorch CNTK	Spark MLlib scikit-learn XGBoost TensorFlow Keras Caffe PyTorch IBM SPSS
Eigene Algorithmen	X		X	X

Tabelle 3.1: Übersicht Plattformbasierte Dienste

Wie man der Tabelle 3.1 entnehmen kann, bieten alle Anbieter ein **Command-line Interface** (CLI) für ihre Plattformen an. Google und Amazon stellen zudem REST APIs zur Verfügung. Amazon, Azure und IBM bieten zusätzlich noch den Zugriff über ein *Jupyter Notebook* an.

Bei den unterstützten Frameworks ist IBM an der Spitze und bietet neben bekannten Open Source Frameworks zusätzlich den hauseigenen *SPSS Modeler* an. Auf Platz zwei befindet sich mit fünf Frameworks Azure, welches mit dem *Microsoft Cognitive Toolkit (CNTK)* ein zusätzliches, von Microsoft entwickeltes, Deep Learning Framework unterstützt. Mit lediglich drei verfügbaren Frameworks liegt Google auf dem letzten Platz. Amazon, Azure und IBM bieten außerdem eigene Algorithmen-Implementationen für Klassifikation und Regression an.

Bezüglich der SDKs liegt Amazon mit acht unterstützten Programmiersprachen vorne. Darauf folgt Google mit sieben Sprachen. Azure und IBM stellen lediglich ein SDK für Python bereit. Falls also für einen Anwendungsfall etwas anderes als Python verwendet wird, fallen Azure und IBM als potenzielle Anbieter weg. Nach einer Umfrage von VisionMobile (Schuermans et al. [2017]) verwenden jedoch 57% von den 2022 befragten ML Entwicklern Python und für 33% ist es die priorisierte Sprache. Daher kann man davon ausgehen, dass in vielen Anwendungsfällen Python Unterstützung ausreicht.

Alle Anbieter stellen durch ihre Frameworks und Algorithmen Lösungen für die gängigsten ML Problematiken (Regression, Klassifizierung) zur Verfügung und decken in dieser Hinsicht denselben Funktionsumfang ab. Es könnten jedoch zusätzliche sekundäre Faktoren wie z.B. Laufzeit, Framework Usability, o.Ä. eine wichtige Rolle in der Entscheidungsfindung spielen. Diese sind jedoch plattformunabhängig und es müssten die Frameworks untereinander verglichen werden.

## 3.2 Vortrainierte ML APIs

Neben den Plattform-Diensten werden zahlreiche ML APIs angeboten, bei denen die Modelle bereits vortrainiert sind. Es werden für die Nutzung keine vorherigen ML Kenntnisse benötigt; man übergibt lediglich die eigenen Daten und erhält die von den Modellen erzeugten Ergebnisse. Die Anbindung erfolgt meist über eine REST Schnittstelle oder SDKs. Durch die bereits vortrainierten und getesteten Modelle entfällt ein großer Arbeitsaufwand, vor allem da keine eigenen Trainings- und Testdatensätze benötigt werden.

Diese vortrainierten Dienste lassen sich zurzeit in drei größere Kategorien einteilen:

1. Sprach- & Textanalyse
2. Bildanalyse
3. Sonstige Dienste

### 3.2.1 Sprach- & Textanalyse

Der Kernbereich der Sprach- und Textanalyse bildet das **Natural Language Processing** (NLP), welches Maschinen ermöglicht mit menschlicher Sprache zu interagieren. NLP besteht unter anderem aus Spracherkennung, Syntaxanalyse und Textgenerierung. Viele dieser Bereiche haben in den vergangenen Jahren deutliche Fortschritte gemacht, nicht zuletzt dank den Erfolgen im Deep Learning (Young et al. [2018]).

In den meisten Fällen wird Spracherkennung zusammen mit NLP genutzt, um die Wünsche von Nutzern nachzuvollziehen. Die bekanntesten Beispiele dieser Technologie sind digitale Sprachassistenten wie Apples *Siri* oder Amazons *Alexa*. In diesem Abschnitt

Funktion / Anbieter	AWS	GCP	Azure	Watson
speech to text	2	119	30	9
text to speech	27	56	51	8
language translator	12	98	62	22
language detection	100	104	66	62
topic extraction	X	X	X	X
entities extraction	X	X	X	X
keywords extraction	X	X	X	X
metadata extraction				X
relations analysis			X	X
syntax analysis	X	X	X	X
sentiment analysis	X	X	X	X
emotion analysis				X
personality analysis				X
tone analysis				X
voice verification			X	
spell check			X	
<b>Total</b>	<b>9</b>	<b>9</b>	<b>13</b>	<b>14</b>

Tabelle 3.2: Funktionsübersicht Sprach- & Textanalyse

werden die unterschiedlichen Sprach- und Textanalyse Tools der Anbieter und deren Funktionsumfang untersucht.

#### Amazon

Amazon bietet für alle sprach- und textrelevanten Aufgaben vier unterschiedliche APIs an, jede mit einem bestimmten Aufgabenbereich. *Amazon Transcribe* dient zur Spracherkennung (Amazon Web Services, Inc. [2018e]). Die dabei erzeugten Audiodateien können dann zur weiteren Analyse verwendet werden. Ein naheliegender Anwendungsfall wäre die Auswertung von Kundentelefonaten. *Transcribe* unterstützt Englisch und Spanisch.

*Amazon Polly* bietet die Möglichkeit zur Sprachsynthese. Dabei wird ein bestehender Text mit einer möglichst menschenähnlichen, computergenerierten Stimme wiedergegeben. Dies ist vor allem hilfreich bei Vorlesegeräten für Sehbehinderte. Polly unterstützt 27 verschiedene Sprachen und bieten für einige Sprachen mehrere Stimmvarianten an (Amazon Web Services, Inc. [2018c]).

Zur Textanalyse wird *Amazon Comprehend* angeboten. Comprehend ist ein NLP Dienst

und bietet unterschiedliche Analysetools an, um in Texten neue Einsichten und Zusammenhänge zu finden (Amazon Web Services, Inc. [2018b]). Aktuell bietet Comprehend folgende Funktionen:

- Schlüsselphrasenextraktion
- Stimmungsanalyse, um die allgemeine Stimmung eines Texts (positiv, negativ, neutral oder gemischt) wiederzugeben
- Syntaxanalyse, mit der man Text in Tokens unterteilen kann, um so Wortgrenzen und Bezeichnungen wie Substantive und Adjektive im Text zu erkennen
- Entitäten Erkennung (Erkennung von Namen, Orten, Unternehmen, etc.)
- Spracherkennung von 100 Sprachen

*Amazon Translate* übersetzt Text aus dem Englischen oder ins Englische. Dabei werden 12 Sprachen unterstützt (Amazon Web Services, Inc. [2018f]).

## Google

Google bietet mit *Cloud Speech-to-Text* ebenfalls einen Spracherkennungsdienst an. Dabei werden insgesamt 119 Sprachen und Dialekte unterstützt (Google LLC [2018e]). Für die Sprachsynthese steht *Cloud Text-to-Speech* zur Verfügung. Dieser Dienst unterstützt 56 verschiedene Sprachen/Dialekte (Google LLC [2018f]). Darunter sind besondere *WaveNet* Stimmen, welche laut Google (van den Oord et al. [2016]) gegenüber anderen Sprachsynthese Technologien von Menschen als authentischer wahrgenommen werden. *WaveNet* erzeugt für die Synthese komplett neue Audiodateien mit Hilfe eines Neuronalen Netzes, welches auf realistische Tonfolgen und Sprachwellen trainiert ist. Diese Technologie wird von Google auch bei seinem *Assistant* und Übersetzungsdienst verwendet.

*Cloud Natural Language* ist das NLP Pendant zu *Amazon Comprehend* und bietet dieselben Funktionalitäten (Google LLC [2018d]).

Mit *Cloud Translation* bietet Google eine Möglichkeit, den bekannten *Google Translate* Übersetzungsdienst in eigene Anwendungen zu integrieren. Dabei liegt Google mit 104 unterstützten Sprachen klar vor der Konkurrenz (Google LLC [2018h]). Im Gegensatz zu *Amazon Transcribe* können bei *Cloud Translation* zudem alle Sprachen auch untereinander übersetzt werden (und nicht nur vom und ins Englische).

#### Azure

Microsoft bietet ebenfalls sprachrelevante Schnittstellen an, welche einen Teilbereich der sogenannten *Cognitive Services* bilden, in denen Microsoft seine KI Dienste zusammenfasst (Microsoft Corp [2018d]). Unter der *Speech API* befinden sich Dienste für Text-to-Speech (51 Sprachen/Dialekte, tlw. mehrere Stimmen zu Auswahl) oder Speech-To-Text (30 Sprachen). Die *Translator Speech API* kombiniert diese beiden Dienste zusammen mit einer Übersetzungsfunktion und bietet damit Speech-to-Speech oder Speech-to-Text Übersetzungen an (Microsoft Corp [2018j]). Azure stellt zudem eine *Speaker Recognition API* zur Verfügung, über die Personen anhand ihrer Stimmen identifiziert und verifiziert werden können (Microsoft Corp [2018h]).

Unter den *Language APIs* findet man die Analyse und Übersetzungstools von Azure. *Text Analytics* bietet NLP für Schlüsselphrasenextraktion, Stimmungsanalyse, Entitäten Erkennung, Beziehungsanalyse zwischen Wörtern/Phrasen und unterstützt die Erkennung von 120 Sprachen (Microsoft Corp [2018a]). Über die *Linguistic Analysis API* ist zudem Tokenisierung möglich (Microsoft Corp [2018f]). Die Übersetzungs-API *Translator Text* unterstützt 65 Sprachen (Microsoft Corp [2018k]). Azure bietet zudem alleinig Rechtschreib- und Grammatikprüfung über die *Bing Spell Check API* an (Microsoft Corp [2018b]).

#### IBM

IBMs Speech-to-Text Service unterstützt neun Sprachen (IBM Corp [2018d]) und mit der Text-to-Speech API können acht verschiedene Sprachen wiedergegeben werden (IBM Corp [2018e]). Der *Language Translator* Dienst bietet die Möglichkeit 21 Sprachen zu übersetzen (IBM Corp [2018g]).

*Natural Language Understanding* analysiert Texte nach Schlüsselwörtern, Entitäten, Relationen und Stimmung und erkennt 62 Sprachen. Der Dienst unterstützt zudem weitere Funktionen, die bei der Konkurrenz nicht verfügbar sind. So können zusätzliche Metadaten extrahiert werden oder zu den gefundenen Schlüsselwörtern und Entitäten verbundene Emotionen abgefragt werden ((IBM Corp [2018b]). Womit sich IBM außerdem von der Konkurrenz abhebt sind die *Personality Insights* und *Tone Analyzer Services*.

Mit *Personality Insights* können komplexere Persönlichkeitsprofile für die Autoren der übergebenen Texte erstellt werden (IBM Corp [2018c]). Diese Profile geben unter ande-

rem Einsicht in Charaktereigenschaften, mögliche Bedürfnisse und Tendenzen zum Konsumverhalten (z.B. bevorzugte Filmgenres). Der *Tone Analyzer* untersucht den Sprachton (analytisch, zuversichtlich, zurückhaltend) und die Emotionen eines Textes (IBM Corp [2018f]).

#### 3.2.2 Bildanalyse

In den vergangenen Jahren gab es immense Fortschritte in der maschinellen Bildbearbeitung. Heutzutage können mit Smartphones in Echtzeit Objekte vor der Kamera erkannt und kategorisiert werden (Redmon [2017]). Ein Großteil dieses Erfolges ist den **Convolutional Neural Networks** (CNN) zuzuschreiben, welche vor allem in der Gesichtserkennung eine wichtige Rolle spielen (Matsugu et al. [2003]). Bei einem Vergleich aller bekannten Bilderkennungsalgorithmen erreichten CNNs die geringste Fehlerquoten (Ciresan et al. [2012]).

Cloudbasierte Services bieten eine leicht zugängliche Schnittstelle, um die Leistung vor-trainierter Netze in eigene Anwendungen zu integrieren.

Primär geht es bei den Bildanalyse-Diensten um die Objekt- und Gesichtserkennung. Beide Funktionen werden von allen vier Anbietern zur Verfügung gestellt. Bei erkannten Gesichtern kann zudem bei allen Anbietern die relative Position im Bild bestimmt werden (**Bounding Box**).

#### Amazon

Amazon *Rekognition* bietet den zweitgrößten Funktionsumfang bei der Bildanalyse (vgl. Tabelle 3.3). Neben Objekten und Gesichtern unterstützt *Rekognition* Texterkennung (**Optical character recognition** (OCR)). Erkannte Gesichter können unter anderem auf Alter, Geschlecht, Emotionen, und Haltung untersucht werden. Zudem kann abgefragt werden, ob es sich bei der erkannten Person um eine prominente Persönlichkeit handelt (Amazon Web Services, Inc. [2018d]).

Bei *Rekognition* können Gesichter mit IDs versehen werden, um so eine Datenbank von bereits analysierten Personen aufzubauen. Dadurch können bekannte Personen auf neuen Bildern wiedererkannt werden. Ähnliche Gesichter auf verschiedenen Bildern können auch



Funktion / Anbieter	AWS	GCP	Azure	Watson
age	X		X	X
facial hair	X		X	
face bounding box	X	X	X	X
celebrity	X		X	
emotion	X	X	X	
glasses	X		X	
eyes open	X			
gender	X		X	X
landmarks	X	X	X	
mouth open	X			
pose	X		X	
picture quality	X	X	X	
smile	X		X	
hair			X	
accessories		X	X	
makeup			X	
text/OCR	X	X	X	
labels/objects	X	X	X	X
logo		X		
<b>Total</b>	<b>15</b>	<b>8</b>	<b>16</b>	<b>4</b>

Tabelle 3.3: Übersicht Bilderkennung

direkt miteinander verglichen werden, um zu prüfen, ob es sich um dieselbe Person handelt. Darüber hinaus bietet *Rekognition* die Möglichkeit, Videos auf Gesichter/Personen/prominente Persönlichkeiten, Objekte und explizite Inhalte zu untersuchen.

#### Google

Googles *Vision* API bietet einen deutlich geringeren Funktionsumfang. Zwar können neben Objekten und Gesichtern zusätzlich Text und Logos erkannt werden, jedoch ist eine detaillierte Analyse eines erkannten Gesichtes nicht möglich (Google LLC [2018j]). Mit der *Video Intelligence* API können Videos auf Objekte, Film-Einstellungen und explizite Inhalte untersucht werden (Google LLC [2018i]).

#### **Azure**

Azure unterstützt mit der *Computer Vision* und der *Face* API die meisten Features. Neben OCR bietet Azure eine detaillierte Gesichtsanalyse, bei der sogar die Haare und das Makeup untersucht werden können (Microsoft Corp [2018e]). Zudem besteht wie bei Amazon *Rekognition* die Option, bereits analysierte Gesichter/Personen zu speichern und verwalten. Im Gegensatz zu Amazon und Google bietet Azure jedoch keine integrierte Videoanalyse.

#### **IBM**

*Watson Visual Recognition* bietet den geringsten Funktionsumfang aller untersuchten Dienste. Neben der Objekt- und Gesichtserkennung besteht lediglich die Möglichkeit, erkannte Personen auf das Alter und Geschlecht zu analysieren (IBM Corp [2018h]). Eine Videoanalyse ist nicht verfügbar.

#### **3.2.3 Sonstige Services**

Neben Bild- und Textanalyse gibt es noch andere vortrainierte ML APIs, die für sehr spezielle Aufgabenbereiche bestimmt sind. So vereinfachen *Azure Bot Service*, *Amazon Lex* und *Watson Assistant* die Erstellung von Bots, welche für die Interaktion mit Menschen bestimmt sind (Microsoft Corp [2018c]; Amazon Web Services, Inc. [2018a]; IBM Corp [2018a]). Mit den *Azure Search* APIs können Anwendungen die Suchfunktionen von Bing integrieren, inklusive Bild-/Videosuche und Autovervollständigung (Microsoft Corp [2018g]). Google bietet mit *Cloud Talent Solution* eine verbesserte Suche für Jobbörsen und optimiert zudem das Einstellungsverfahren für Unternehmen (Google LLC [2018g]).

### **3.3 (Semi-)Automatisierte Dienste**

Die letzte Kategorie der ML Services bilden die (semi-)automatisierten Dienste. Diese sollen den Aufbau von eigenen Modellen stark vereinfachen bzw. vollständig automatisieren. Man übergibt die eigenen Trainingsdaten und die Dienste wählen dann die passenden Algorithmen aus, um die Modelle zu trainieren.

### 3 Angebotsübersicht

---

	Amazon ML	Google AutoML	Azure ML Studio	IBM Watson Studio
Regression	X		X	X
Binary Classification	X		X	X
Multiclass Classification	X		X	X
Clustering			X	
Anomaly Detection			X	
Recommendation			X	
Statistical Functions			X	
NLP		X	X	
Computer Vision		X	X	

Tabelle 3.4: Übersicht autom. Dienste

Amazon und IBM bieten dabei die höchste Automatisierung an. Man wählt lediglich das gewünschte Verfahren (Regression, Binäre/Mehrklassen-Klassifizierung) aus und den Rest übernimmt der Dienst. IBM unterstützt zusätzlich einen manuellen Modus, bei dem 12 Algorithmen zur Auswahl stehen.

*Azure Machine Learning Studio* stellt die meisten Funktionen zur Verfügung. Neben Regression und Klassifizierung können noch weitere Problemstellungen angegangen werden, wie z.B. Clustering oder Anomalie Erkennung. Insgesamt stehen über 30 verschiedene Algorithmen zur Auswahl (Microsoft Corp [2018i]). Zusätzlich besteht noch die Möglichkeit zur Daten-Exploration und Vorverarbeitung.

Google bot mit der *Cloud Prediction API* ebenfalls eine automatisierte Lösung für Regression und Klassifizierung an, stellte diese aber am 30. April 2018 ein. Mit *Cloud AutoML* ermöglicht Google jedoch seit neuestem das automatische Trainieren von Modellen zur Bild- und Textanalyse (Google LLC [2018b]). Für *AutoML Vision* stellt Google sogar ein internes Team zur Verfügung, das die kundenspezifischen, ungekennzeichneten Bilder nach den Instruktionen des Kunden kategorisiert, um diese für das Training weiterzuverwenden (Google LLC [2018c]).

## 4 Evaluierung

Für die Evaluierung der Dienste sind verschiedene Ansätze möglich, je nachdem welche Dienstkategorie untersucht werden soll. Bei der Betrachtung der Plattformdienste könnte zu einem die darunter liegende Hardware verglichen werden, da diese einen großen Einfluss auf die Performance von Modellen haben kann (Jawandhiya [2018]). Weiterhin wäre auch eine Betrachtung einer Laufzeitanalyse beim Training der Modelle mit dem selben Framework möglich. Bei der Qualität der Modellvorhersagen können jedoch vermutlich keine bis geringe Unterschiede festgestellt werden, wenn dasselbe Framework genutzt wird; und eine Analyse mit unterschiedlichen Frameworks hätte keine wirkliche Aussagekraft für die Plattformdienste selbst. Der Vergleich zwischen den (semi-)automatisierte Diensten fällt bereits dadurch schwer, dass es kein gemeinsames Feature gibt, welches von allen Anbietern unterstützt wird (vgl. Tabelle 3.4). Daher wird der Fokus dieser Evaluierung auf die vortrainierten ML APIs gesetzt. In dieser Kategorie stehen die Anbieter in starker Konkurrenz und bieten ähnliche Funktionen in der Text- und Bildanalyse an. Da die jeweiligen Modelle von den Anbietern selbst aufgebaut und trainiert wurden, ist es für potentielle Nutzer interessant, die Qualität dieser Modelle zu untersuchen. Hierfür wurden im Rahmen dieser Arbeit jeweils für die Bild- und Textanalyse Tests entworfen und Bewertungskriterien festgelegt, um die Anbieter untereinander zu bewerten.

### 4.1 Testumgebung

Die Tests zur Bild- und Textanalyse wurden unter MacOS 10.13.6 und mittels Python 3.7.0 entwickelt. Zudem wurden folgende anbieterspezifischen Bibliotheken/SDKs genutzt:

Azure	cognitive_face 1.4.2
AWS	boto3 1.7.72
GCP	google-cloud-vision 0.33.0, google-cloud-language 1.0.2, googleapis-common-protos 1.5.3
IBM	watson-developer-cloud 1.5.0

## 4.2 Bildanalyse

Zur Bewertung von Bildanalyseediensten wird die Leistung der Gesichtserkennung gemessen. Das Gesicht ist eines der wichtigsten biometrischen Merkmale des Menschen und dient unter anderem zur Identifikation und Verifikation. In einigen Flughäfen in den USA sind bereits Systeme zur Identifizierung von Passbetrügern oder Personen auf Beobachtungslisten im Einsatz (Fingas [2017]). In Deutschland werden aktuell solche Systeme im öffentlichen Raum ebenfalls getestet (Krempl [2018]). Mittlerweile lassen sich sogar Smartphones mit Hilfe von Gesichtserkennung entsperren und auch soziale Netzwerke wie Facebook nutzen Gesichtserkennung, um ihre Dienste interaktiver zu gestalten. Ziel des Tests ist es herauszufinden, wie gut die unterschiedlichen Dienste Gesichter in Bildern erkennen.

### 4.2.1 Datensatz



Abbildung 4.1: Beispielbilder aus dem LFW Datensatz

Für den Test wurde ein Datensatz aus Bildern mit und ohne Gesichtern erstellt. Der Datensatz besteht aus insgesamt 298 Bildern im JPEG-Format. Auf 198 davon sind menschliche Gesichter abgebildet. Diese stammen vom Datensatz *LFW* (Huang et al. [2007]), welches 13,000 Bilder von 1680 Personen aus dem Internet beinhaltet. Für den Test wurden alle Bilder der ersten 100 Personen nach alphabetischer Sortierung ausgewählt. Die restlichen 100 Bilder stammen aus dem Datensatz *Caltech 101* (Fei-Fei et al. [2006]). Dieser beinhaltet 9146 Bilder aus 101 unterschiedlichen Objekt Kategorien (Uhren, Ameisen, Pianos, etc.). Hier wurde jeweils das erste Bild jeder einzelnen Kategorie zum Test-Datensatz hinzugefügt, ausgenommen der Kategorie *Faces*, welches als einzige Bilder mit menschlichen Gesichtern beinhaltet. Eine manuell angelegte CSV-Datei hält alle Dateinamen fest, auf denen Gesichter abgebildet sind und dient als *Ground Truth*.

Gegen diese werden die Ergebnisse der Dienstanalysen bei der Auswertung geprüft. Zusätzlich sind dort die jeweiligen Geschlechter der Personen vermerkt, die manuell ermittelt wurden.

### 4.2.2 Aufbau

Das Hauptziel des Tests ist es, menschliche Gesichter in einer Sammlung von Bildern zu erkennen. Dafür wurden zunächst bei Azure, GCP und IBM Cloud über die Webinterfaces die jeweiligen Cloud-Dienste erstellt. Bei allen drei Anbietern wurde die kostenlos verfügbare Stufe ausgewählt. Diese limitiert die Anzahl der kostenfreien Aufrufe pro Monat (Azure: 30000/Monat, GCP: 1000/Monat, IBM: 1000/Monat). Bei Azure gibt es zusätzlich eine Einschränkung von 20 Aufrufen pro Minute. Bei AWS muss der Dienst nicht explizit angelegt werden, es muss lediglich die Erlaubnis zur Nutzung für den jeweiligen Anwender im *AWS Identity- and Access Management* erteilt werden. AWS bietet bis zu 5000 kostenlose Aufrufe pro Monat.

Über Service-Endpunkte können bei AWS und Azure die gewünschte Region des Servers ausgewählt werden. IBM bietet seinen Bilderkennungsdienst aktuell nur in den USA (Dallas) an. Bei Azure wurde die Region *EU West* (Serverstandort: Niederlande) ausgewählt; für AWS wurde ebenfalls *EU West* (Serverstandort: Irland) genutzt. Google bietet keine Auswahl bezüglich der Region mit dem Hinweis: „*Der Service wird weltweit angeboten und hängt nicht von einem bestimmten Standort ab.*“ (Google LLC [2018a])

IBM und Google unterstützen für ihre Dienste verschiedene Versionen, die für die API-Anfragen ausgewählt werden können. Google bietet neben seiner stabilen Version **v1** zwei Beta Versionen, mit aktuelleren Features. IBM nutzt die *Foursquare Versionierung* (Foursquare Labs, INC [2018]) und stellt eine neue API Version zur Verfügung, sobald durch Änderungen keine Rückwärtskompatibilität mehr besteht. Aktuell existiert jedoch nur eine Version. Für diese Arbeit wurde bei Google die stabile Version **v1** und für IBM die einzig verfügbare Version **2018-03-19** verwendet.

Nach der Erstellung der Dienste wurden die jeweiligen Zugangsdaten generiert. IBM, AWS und Azure nutzen hierfür API Keys. Die Keys werden über die Webinterfaces automatisch generiert und müssen bei der Nutzung des Dienstes mit übergeben werden. Diese Keys bestehen aus unterschiedlich langen Zeichenketten (IBM:44, Azure:32, AWS:40). Google nutzt für den Zugriff auf die Dienste sogenannte Service Accounts.

Diese sind den jeweiligen API nutzenden Applikationen zugewiesen. Jeder Service Account ist mit einem oder mehreren Schlüsselpaaren assoziiert, die über GCP erstellt und verwaltet werden; diese Methode ähnelt einer Public-Key-Infrastruktur. Zur Erzeugung der Keys kann entweder ein nicht spezifizierter Algorithmus von Google oder alternativ RSA (Rivest et al. [1978]) 1024 oder 2048 Bits genutzt werden. Nach der Generierung eines neuen Schlüsselpaares im Webinterface wird der private Schlüssel auf dem Host der Anwendung abgespeichert und dient dieser zur Authentifizierung.

Für die Analyse der Bilder und um deren Annotation (d.h. die Ergebnisse der Dienstanalyse) zu speichern wurde ein Python Skript entwickelt, welches die bereitgestellten Bibliotheken der Anbieter nutzt. Die API-Key von Azure und IBM wurden über Umgebungsvariablen des Betriebssystems abgerufen.

Für GCP musste die Umgebungsvariable `GOOGLE_APPLICATION_CREDENTIALS` mit dem Pfad zu dem gespeicherten privaten Key gesetzt werden. Die AWS Boto3 Bibliothek bezieht den Key über eine Zugangsdaten Datei (`~/.aws/credentials`).

```
def aws_annotate(aws_client , file ):
    with open( file , 'rb' ) as img:
        content = image.read()
        annotation = aws_client.detect_faces( # API Aufruf
            Image={'Bytes': content},
            Attributes=['ALL'] )
    return annotation
```

Listing 4.1: AWS Annotationfunktion

Dem Skript werden zwei Parameter übergeben. Zum einen der Pfad zu den hinterlegten Bildern und der Name des Anbieters, der die Bilder annotieren soll. Anhand des Namens wird der anbieterspezifische Client für den Cloud-Service konfiguriert und die jeweilige Annotierungsfunktion ausgewählt, welche den Cloud-Dienst des jeweiligen Anbieters aufruft. Nach dieser initialen Konfiguration iteriert das Skript über alle Bilder im Datensatz und lässt jedes Bild einzeln von dem Cloud Service analysieren. Azure, AWS und IBM bieten bei ihren APIs explizite Methoden für die Gesichtserkennung; Google bietet lediglich eine allgemeine Annotation Funktion, welche alle Bildanalysedienste vereint, daher muss beim Methoden Aufruf zusätzlich der Feature-Parameter `FACE_DETECTION` übergeben werden. Bei den Serviceaufrufen für AWS und Azure können neben dem Bild

noch Gesichts-Features übergeben werden, auf die das Bild zusätzlich zur allgemeinen Gesichtserkennung analysiert werden soll. AWS bietet dafür zwei Kategorien: DEFAULT und ALL, wobei DEFAULT die Analyse auf Bounding Box, Zuversicht, Stellung, Qualität, und Gesichtsmerkmale einschränkt. Azure bietet eine feinere Granularität, bei der jedes Feature einzeln ausgewählt werden kann. Für diesen Test wurden die Bilder bei Amazon auf alle Features (ALL) und bei Azure zusätzlich auf das Geschlecht analysiert.

Mit Ausnahme von Google senden alle Anbieter ihre Annotationen im JSON Format zurück. Google nutzt bei der Kommunikation zwischen Client und Cloud Service die hausintern entwickelten *protocol buffers*, welche eine sprach- und plattformneutrale Methode zur Serialisierung strukturierter Daten bietet. Bei dem Serviceaufruf selbst ist dies nicht offensichtlich, da das SDK die direkte Kommunikation mit dem Dienst abstrahiert. Zur Weiterverwendung der zurückgeschickten Annotation muss jedoch die Protocol Buffers Bibliothek für Python eingesetzt werden, um die Antwort in ein normales JSON Objekt umzuwandeln.

Bei jedem Iterationsschritt werden der Dateiname und die Zugehörige JSON Annotation in einem Array gesammelt.

```
{
  "image": "data/images/Al_Gore_0005.jpg",
  "annotation": [
    {
      "faceId": "2f0d241f-997b-42bb-9bed-5a9ff9ecf63b",
      "faceRectangle": {
        "top": 85,
        "left": 77,
        "width": 95,
        "height": 95
      },
      "faceAttributes": {
        "gender": "male"
      }
    }
  ]
}
```

Listing 4.2: Beispiel Azure Annotierung



Nachdem die Analyse aller Bilder beendet ist, werden die Ergebnisse als JSON Datei abgespeichert. Da der Azure Dienst in der kostenlosen Stufe eine zusätzliche Einschränkung von 20 Aufrufen/Minute hat, musste die entsprechende Annotierungsfunktion mit einer Schleife erweitert werden, welche den von der API erzeugten Fehler zur Überschreitung des Limits abfängt und das Programm pausiert, bis wieder Anfragen möglich sind. Die von dem Skript erzeugten JSON Dateien wurden anschließend zur Auswertung weiterverarbeitet.

Neben den Annotationen wurde die Laufzeit der Dienste festgehalten. Dafür wurden die Zeit zwischen den API Aufruf und der Antwortzeit jedes Bildes mit der Python `time` Bibliothek gemessen und pro Anbieter aufsummiert.

### 4.2.3 Bewertungskriterien

Bei der Gesichtserkennung handelt es sich um ein klassisches binäres Klassifizierungsproblem. Die Grundlage für die Bewertung eines solchen Klassifikators bildet die Wahrheitsmatrix. Diese zeigt die Häufigkeit der richtigen oder falschen Klassifikationen. Die möglichen Fälle für diesen Test sind::

1. Richtig positiv (true positive, TP): Ein Bild enthält ein Gesicht, und der Dienst hat dies richtig erkannt
2. Falsch positiv (false positive, FP): Ein Bild enthält kein Gesicht, aber der Dienst hat fälschlicherweise eines erkannt
3. Richtig negativ (true negative, TN): Ein Bild enthält kein Gesicht, und der Dienst hat dies richtig erkannt
4. Falsch negativ (false negative, FN): Ein Bild enthält ein Gesicht, aber der Dienst hat dieses nicht erkannt

Mit Hilfe der Wahrheitsmatrix können unterschiedliche Metriken berechnet werden, um den Klassifikator zu bewerten. Die bekannteste und meist genutzte Metrik (Hossin and Sulaiman [2015]) ist Accuracy.

### Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy zeigt den Anteil der insgesamt vom Modell korrekt getroffenen Vorhersagen an; in diesem Fall alle richtigen mit und ohne Gesicht annotierten Bilder. Neben der Accuracy werden für diesen Test Precision und Recall untersucht, welche zusätzliche Kenntnisse liefern und für spezielle Anforderungen von Bedeutung sind.

### Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision gibt den Prozentanteil aller „mit Gesicht“ klassifizierten Bilder wieder, bei denen die Vorhersage richtig war. Eine höhere Precision ist für Anwendungen empfehlenswert, die möglichst wenig falsche positive Vorhersagen treffen soll.

### Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall ist der Anteil der richtigen Bilder mit Gesicht, die auch als solche klassifiziert wurden. Falls eine Anwendung möglichst wenig falsche negative Vorhersagen treffen soll, ist ein hoher Recall zu empfehlen.

Die Laufzeiten der Bildanalysen werden nicht in die Auswertung miteinbezogen. Durch die regionalen Unterschiede kann die Latenz eine prägnante Auswirkung auf die Zeitmessung haben. Zudem hat die unterschiedliche Anzahl an Bildanalyse-Features der Anbieter einen Einfluss auf die Laufzeiten, wodurch ein Vergleich wenig Aussagekraft hat.

Neben der Gesichtserkennung wurden für AWS, Azure und IBM zusätzlich die Accuracy der Geschlechtserkennung untersucht, welche ein Teil der Gesichtserkennung ist (GCP bietet diese Feature nicht an). Für die Bewertung wurden jeweils die vom Anbieter korrekt erkannten Gesichtsbilder genutzt (dh. die jeweiligen true positive). Die möglichen Fälle für die Geschlechtserkennung sind:

1. TP: Die Person ist weiblich und wurde vom Dienst so kategorisiert

2. FP: Die Person ist männlich, wurde jedoch als weiblich kategorisiert
3. NP: Die Person ist männlich und wurde vom Dienst so kategorisiert
4. FN: Die Person ist weiblich, wurde jedoch als männlich kategorisiert

#### 4.2.4 Auswertung

Für die Auswertung wurden die Annotationen der Anbieter mit der *Ground Truth* verglichen, um so die Anzahl der richtigen und falschen Vorhersagen zu ermitteln (s. Tabelle 4.1).

	AWS	Azure	GCP	IBM
TP	198	197	198	197
FP	1	0	0	4
TN	99	100	100	96
FN	0	1	0	1

Tabelle 4.1: Wahrheitsmatrix Gesichtserkennung

Anhand dieser Werte wurden dann die Metriken für jeden Anbieter errechnet:

	AWS	Azure	GCP	IBM
Accuracy	0,996	0,996	1	0,983
Precision	0,994	1	1	0,98
Recall	1	0,994	1	0,994

Tabelle 4.2: Gesichtserkennung Ergebnisse

Google liegt bei allen Bewertungskriterien der Gesichtserkennung vorne und hat als einziger Anbieter alle Bilder korrekt kategorisiert. Amazon und Azure liegen bei der Accuracy gleichauf, jedoch bietet Azures Modell eine höhere Precision, dass von Amazon hingegen einen höheren Recall. Azure ist daher eher für Anwendungsfälle geeignet, bei denen es wichtig ist, dass positiv annotierte Bilder auch tatsächlich ein menschliches Gesicht enthalten; Amazon bei Anwendungen, die möglichst viele Bilder mit Gesicht erkennen soll. Das Modell von IBM annotierte die Bilder am häufigsten falsch und liegt an letzter Stelle.

Bei der Geschlechtererkennung (Tabelle 4.3) konnte Azure bei allen Personen das Geschlecht richtig ermitteln. AWS und IBM hatten bei einigen Personen Probleme, wobei AWS bei einem Bild zusätzlich falsch lag. Jedoch hatte AWS im Gegensatz zu Azure

	AWS	Azure	IBM
TP	19	20	19
FP	3	0	2
TN	175	177	175
FN	1	0	1
Accuracy	0,98	1	0.985

Tabelle 4.3: Geschlechtserkennung Ergebnisse

und IBM bei der Gesichtserkennung ein Bild mehr als korrekt annotiert und musste dementsprechend dieses ebenfalls bei der Geschlechtserkennung analysieren.

### 4.3 Textanalyse

Der zweite Teil der Evaluierung wird die Qualität der Sentimentanalyse untersuchen. Sentimentanalyse verwendet *Natural language processing*, um die Stimmung eines Texts zu extrahieren. Es wird analysiert, ob der Text im Allgemeinen als positiv oder negativ wahrgenommen wird oder eher neutral/gemischt ist. Da das manuelle Klassifizieren von Social Media Einträgen oder Produktrezessionen auf ihr Sentiment als zeitintensive und mühsame Arbeit gilt, sind ML gestützte Dienste in diesem Bereich sehr gefragt.

Für Unternehmen bietet sich so ein effektives Mittel, um die Meinungen von Verbrauchern bezüglich Marken, Produkten oder Events zu ermitteln.

#### 4.3.1 Datensatz

Für die Sentimentanalyse wurde ein anonymisierter Datensatz eines E-Commerce Shops für Frauenkleidung genutzt (Brooks [2018]). Dieser beinhaltet ca. 23.5 Tsd. Kundenbewertungen für die Produkte. Die Bewertungen bestehen unter anderem aus einer eindeutigen Bewertungs-ID, einer Produktbewertung von eins (negativ) bis fünf (positiv) und einer optionalen, englischsprachigen Rezension.

ID	Produkt-ID	Alter	Titel	Rezension	Produktbewertung
64	862	43	Love this top	Loved this top and was really happy to find it on sale!	5

Tabelle 4.4: Beispiel Bewertung

Für den Test wurden zunächst alle Bewertungen ohne Rezension entfernt. Anschließend wurden alle Bewertungen mit einer Rezensionslänge von unter 15 Zeichen entfernt, da diese auf Grund der Kürze nicht vom IBM Dienst analysiert werden können.

Von den verbleibenden Bewertungen wurden die ersten 5000 (aufsteigend sortiert nach der ID) verwendet.

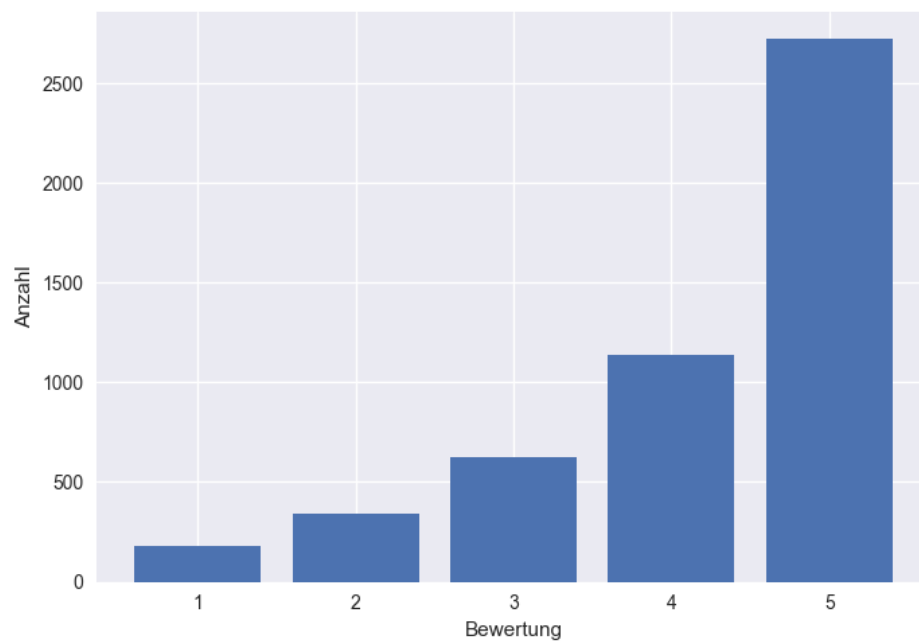


Abbildung 4.2: Verteilung der Produktbewertungen

### 4.3.2 Aufbau

Bei der Sentimentanalyse soll die Fähigkeit der Services geprüft werden, anhand der Rezension das Sentiment einer Bewertung zu bestimmen. Die numerische Produktbewertung dient dabei als *Ground Truth*.

Wie bei der Gesichtserkennung wurden zunächst die jeweiligen Cloud-Dienste über die Webinterfaces erstellt (erneut kostenlose Stufen). Die Authentifizierung für die Dienste

bei Amazon, Azure und Google erfolgte analog zu der Gesichtserkennung. IBM verwendet für den Textanalyse Dienst eine Username und Passwort basierte Authentifizierung anstelle eines API Keys.

Die Dienste bei AWS und Azure wurden wieder für die Region *EU West* konfiguriert. Für IBM wurde der Service Endpunkt für Frankfurt verwendet. Google bietet auch für den NLP Dienst keine Auswahl bei der Region, mit demselben Vermerk wie bei der Bildanalyse. Für die Versionierung wurde bei Google wieder die stabile Version **v1** der API genutzt, bei IBM die Version **2018-03-16**.

Das verwendete Pythonskript zur Sentimentbestimmung ist ähnlich aufgebaut wie bei der Bildanalyse. Dem Skript werden der gewünschte Anbieter und der Datensatz als CSV Datei übergeben. Das Skript iteriert über die Bewertungen und lässt die dazugehörige Rezension von den jeweiligen Diensten analysieren. Die Antwortformate sind identisch zur Bildanalyse (GCP: protocol buffers, Rest: JSON). Im Gegensatz zur Bildanalyse bietet Azure jedoch kein Python SDK für die Textanalyse an. Die API Anfragen wurden daher über die HTTP-Bibliothek `Request` zusammengestellt.

Bei den Anfragen für AWS und Azure musste neben dem zu analysierenden Text zudem die Sprache über einen Sprachcode angegeben werden. IBM und GCP erkennen die Textsprache automatisch.

Bei den Rückgabewerten der Sentimentanalyse unterscheiden sich die Anbieter. Google und IBM bewerten anhand eines Wertes zwischen -1 (negatives Sentiment) und 1 (positives Sentiment). Bei Azure liegt der Wert zwischen 0 (negativ) und 1 (positiv). AWS bewertet das Sentiment lediglich anhand von Strings: `POSITIVE`, `NEGATIVE`, `NEUTRAL` und `MIXED`. Neben dieser Bewertung wird bei AWS zudem ein Wert zwischen 0 und 1 für jedes Label mitgeschickt, welches die Höhe der Zuversicht des Sentiments wiedergibt.

```
{
  "document_sentiment": {
    "magnitude": 0.800000011920929,
    "score": 0.800000011920929
  },
  "language": "en",
  "sentences": [
    {
      "text": {
        "content": "Beautiful colors",
        "begin_offset": -1
      },
      "sentiment": {
        "magnitude": 0.800000011920929,
        "score": 0.800000011920929
      }
    }
  ]
}
```

Listing 4.3: Beispiel Antwort GCP Sentimentanalyse

IBM und GCP geben neben dem Sentiment zusätzlich die erkannte Sprache zurück. Außerdem sendet IBM ein Label für das Sentiment, welches jedoch lediglich eine grobe Unterteilung des numerischen Wertes widerspiegelt (Sentiment  $<0$ : negativ,  $>0$ : positiv,  $=0$ : neutral). Googles Antwort enthält zusätzlich einen `magnitude` Wert (zwischen  $[0, +\infty]$ ), welcher die Stärke des Sentiments darstellt, unabhängig von der Sentimentbewertung. Außerdem unterteilt Google automatisch den übergebenen Text in einzelne Sätze, analysiert diese auf jeweils eigene Sentiment und `magnitude` Werte und sendet diese in der Antwort zurück.

Für die Auswertung wurden die Ergebnisse der Sentimentanalysen mit der dazugehörigen Bewertungs-ID als JSON Datei abgespeichert. Zusätzlich wurde die Zeit zwischen API Aufruf und Antwort gemessen.

### 4.3.3 Bewertungskriterien

Durch die unterschiedlichen Skalen der Anbieter fällt ein direkter Vergleich des Sentiments pro Bewertung schwer. Für einen Vergleich zwischen Azure, IBM und Google bestünde die Möglichkeit, die Skalen zu normieren. Da AWS jedoch lediglich mit Labeln

arbeitet, fällt diese Option als Bewertungsgrundlage für den Vergleich aller Anbieter weg.

Ein weiterer Entwurf sah vor, die Skalen von Azure, IBM und GCP in Bereiche aufzuteilen und diesen ein Label zuzuweisen (z.B. -1 bis -0.5: negativ, -0.5 bis 0.5: neutral, 0.5 bis 1: positiv). Bei dieser Methode besteht jedoch die Problematik, wie und ob solche Grenzen sinnvoll gesetzt werden können.

Damit alle vier Anbieter methodisch bewertet werden können, wurde daher das Problem auf eine Binäre-Klassifizierung reduziert. Die Bewertungen wurden in zwei Kategorien unterteilt: alle Bewertungen  $< 3$  gelten als negativ, alle Bewertungen  $> 3$  als positiv. Produktbewertungen von 3 werden für die Sentiment Bewertung nicht berücksichtigt. Die Sentiment Werte bei IBM und Google werden bei  $< 0$  als negativ und bei  $> 0$  als positiv eingestuft. Bei Azure sind alle Werte  $< 0.5$  negativ und  $> 0.5$  positiv. Für AWS gelten die Label NEGATIVE entsprechend als negative und POSITIVE als positive Bewertung.

	Negatives Sentiment Bewertung $< 3$	Positives Sentiment Bewertung $> 3$
AWS	NEGATIVE	POSITIVE
Azure	$< 0,5$	$> 0,5$
GCP	$< 0$	$> 0$
IBM	$< 0$	$> 0$

Tabelle 4.5: Sentiment-Kriterien für die Kategorisierung

Bei dieser Methode werden also alle Rezensionen mit einer Bewertung von 3 ignoriert und alle Sentiment Bewertungen von 0 (IBM, GCP), 0.5 (Azure) und MIXED/NEUTRAL (AWS) als nicht korrekt klassifiziert angesehen, da das Sentiment vom Service nicht eindeutig als positiv (Produktbewertung  $> 3$ ) oder negativ (Produktbewertung  $< 3$ ) eingestuft werden konnte. Anhand dieser Bewertungsmethode können die Anbieter über die Accuracy ihrer Bewertungen sinnvoll verglichen werden. Neben der Accuracy wurden zudem die Laufzeiten miteinander verglichen.

### 4.3.4 Auswertung

Um einen ersten Einblick in die Resultate der Sentimentanalyse zu gewinnen, wurden für Google, IBM und Azure Violin-Plots erstellt. Diese zeigen die Verteilung der Sentimentwerte auf die jeweilige Produktbewertung. Bei der Erstellung der Plots fiel auf, dass bei



den Analysen von Google teilweise die Sentimentbewertungen fehlten. Die Rezensionen wurden jedoch fehlerfrei analysiert, die entsprechenden Resultate beinhalten bis auf den Sentimentwert alle anderen Ergebniswerte (‘magnitude‘ Wert, erkannte Sprache, erkannte Sätze). Es sind sogar die Sentimentwerte der einzelnen Sätze vorhanden, lediglich die Gesamtbewertung fehlt. Da diese Gesamtbewertung nicht trivial durch die Einzelbewertungen der Sätze aggregiert werden kann, wurden die betroffenen Bewertungen für alle Anbieter aus dem Bewertungsverfahren entfernt.

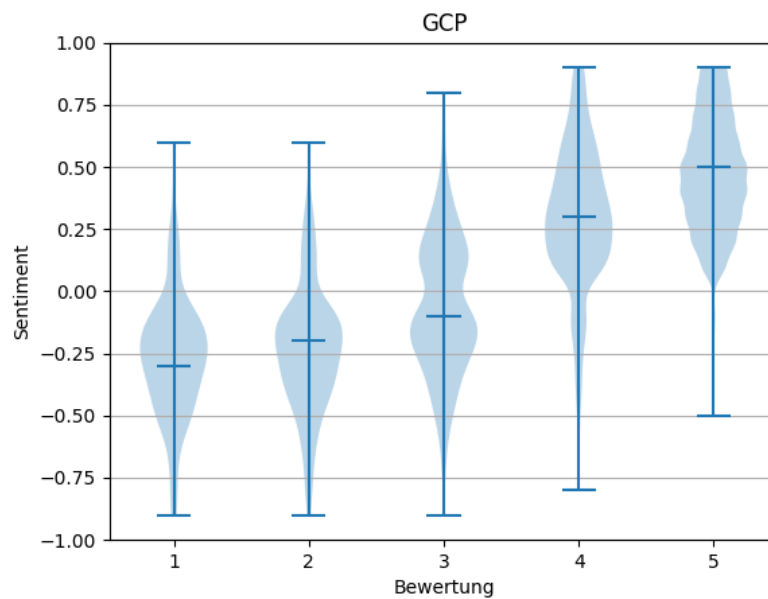


Abbildung 4.3: GCP Sentiment Verteilung

Bei der Betrachtung der Violin-Plots fällt auf, dass sowohl Azure als auch IBM bei fast allen Bewertungen Ausreißer über die jeweilig gesamten Skalen aufzeigen, was auf eine sehr große Unsicherheit in der Beurteilung hindeutet. Bei Azure sind zudem kaum Bewertungen im Mittelbereich der Skala (um 0.5) zu finden, das Modell tendiert eher zu extremen Werten.

Bei Googles Ergebnissen kann man anhand der Verteilung einen verhältnismäßigen Anstieg des Sentimentwertes zur Produktbewertung erkennen, so wie man ihn erwarten würde. Wenn man jedoch die Durchschnittswerte über alle Bewertungen hinweg betrachtet, fällt auf, dass lediglich ein Bruchteil der möglichen Skala genutzt wird (ca. -0.25 bis 0.5). Dadurch ist das Modell bei seinen Bewertungen nicht so aussagekräftig wie bei anderen Anbietern.

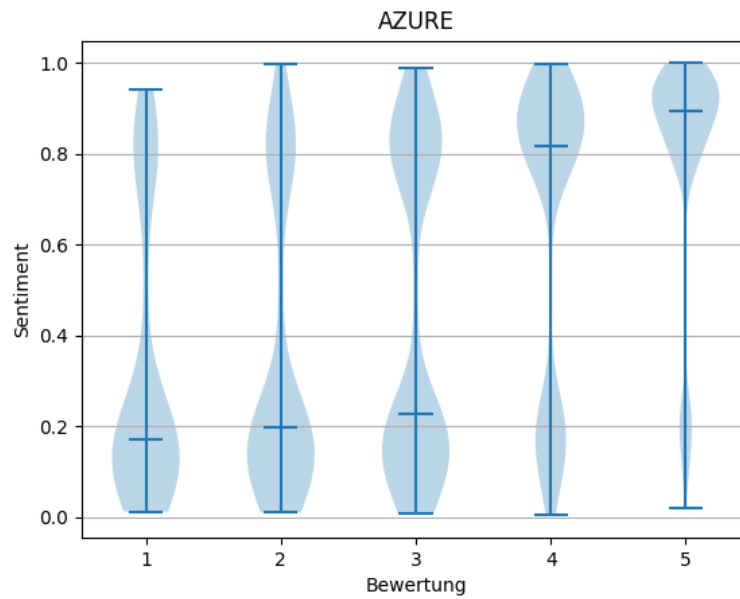


Abbildung 4.4: Azure Sentiment Verteilung

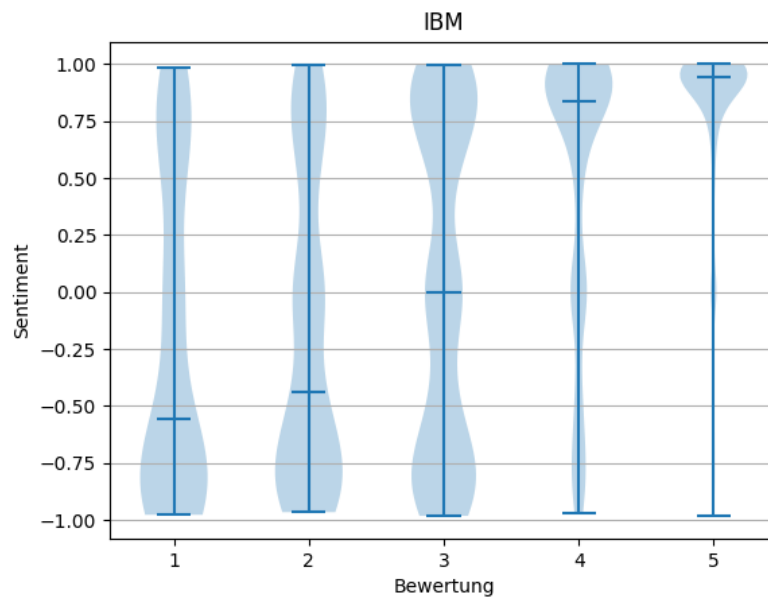


Abbildung 4.5: IBM Sentiment Verteilung

Neben den Violin-Plots wurden für die drei Anbieter Histogramme erstellt, welche die

Sentimentverteilungen für die Rezensionen mit einer Bewertung von 3 zeigen. Alle drei Anbieter liegen hier mit ihrer Durchschnittsbewertung dort, wo man es anhand ihrer Skalen für die 3er Bewertungen erwarten würde: Azure Mitte 0.5,  $\sigma$  0.44; IBM Mitte 0,  $\sigma$  0.03; GCP Mitte 0,  $\sigma$  -0.06. Jedoch weist lediglich Google eine Normalverteilung für diese Bewertungen auf.

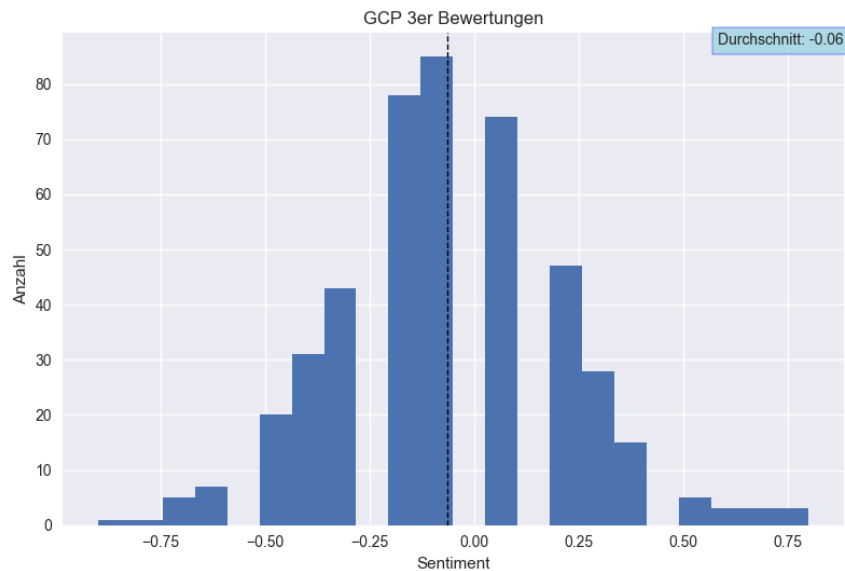


Abbildung 4.6: Sentiment Verteilung für 3er Rezensionen bei GCP

Bei Azure und IBM liegt der Durchschnitt zwar im erwarteten Bereich, jedoch ergibt sich dieser durch das Mittel an zu negativ oder zu positiv Bewerteten Rezensionen. Vor allem bei Azure fällt auf, wie wenig der 3er Rezensionen im Mittel der Skala liegen. IBM und Azure eignen sich daher nicht für Textanalysen mit neutralem/gemischtem Sentiment.

Zur Einsicht der AWS Ergebnisse wurde ein Säulendiagramm (Abbildung 4.9) erstellt, welches die Verteilung der Labels auf die jeweiligen Produktbewertungen zeigt. Dabei ist gut zu sehen, dass das Modell bei den 1er und 5er Rezensionen das Sentiment richtig eingestuft hat. Jedoch zeigt auch das AWS Modell Schwächen bei gemischten/neutralen Rezensionen. Von den gut 600 3er Rezensionen wurden jeweils über ein Drittel als positiv oder negativ kategorisiert.

Um die Analysen nun nach ihrer Accuracy bewerten zu können, wurde zunächst eine Liste aller relevanten Bewertungen zusammengestellt. Von den ursprünglichen 5000 Bewertungen wurden zunächst alle 604 Bewertungen entfernt, welche keinen Sentiment Wert

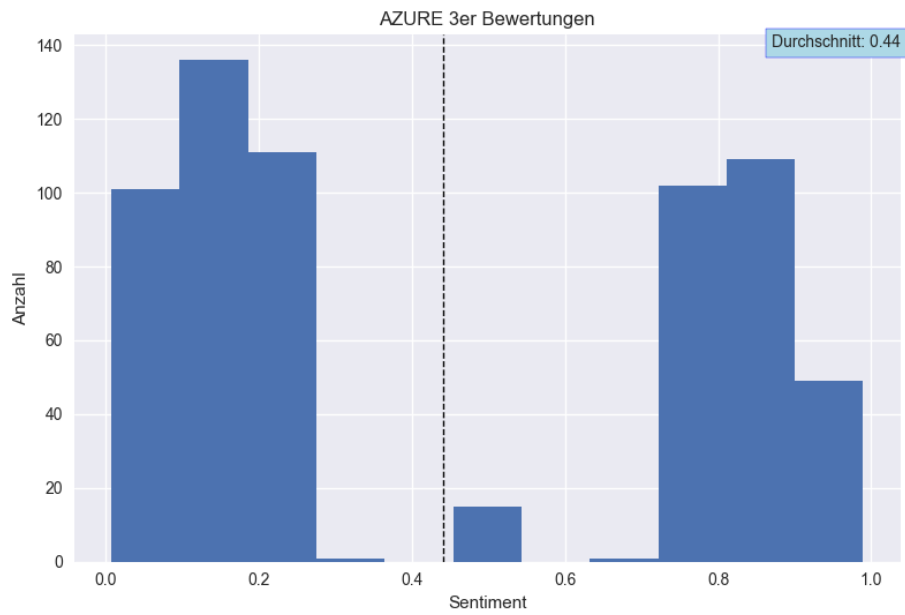


Abbildung 4.7: Sentiment Verteilung für 3er Rezensionen bei Azure

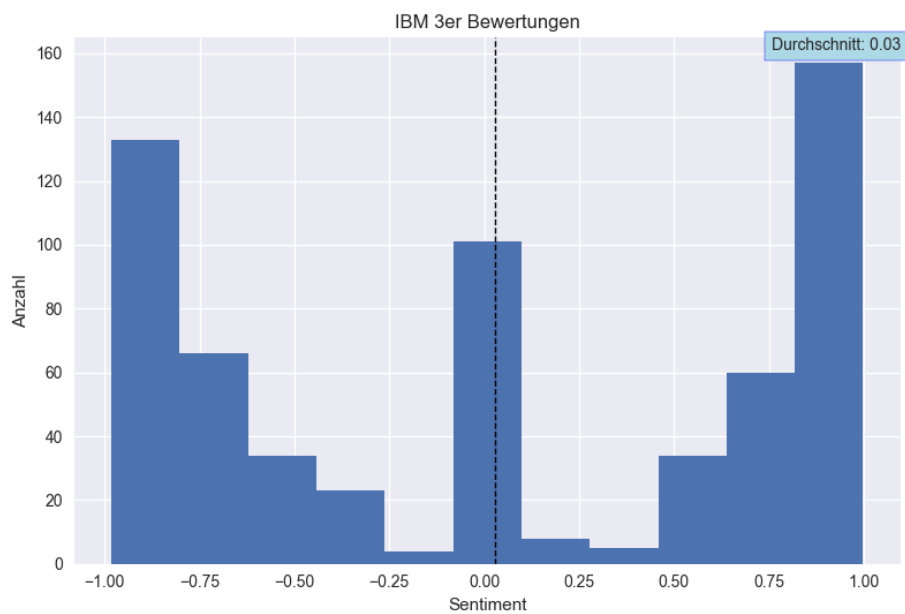


Abbildung 4.8: Sentiment Verteilung für 3er Rezensionen bei IBM

von Google erhalten haben. Danach wurden die übrig gebliebenen 449 3er Bewertungen gestrichen. Damit bleiben für die Endauswertung 3947 Bewertungen.

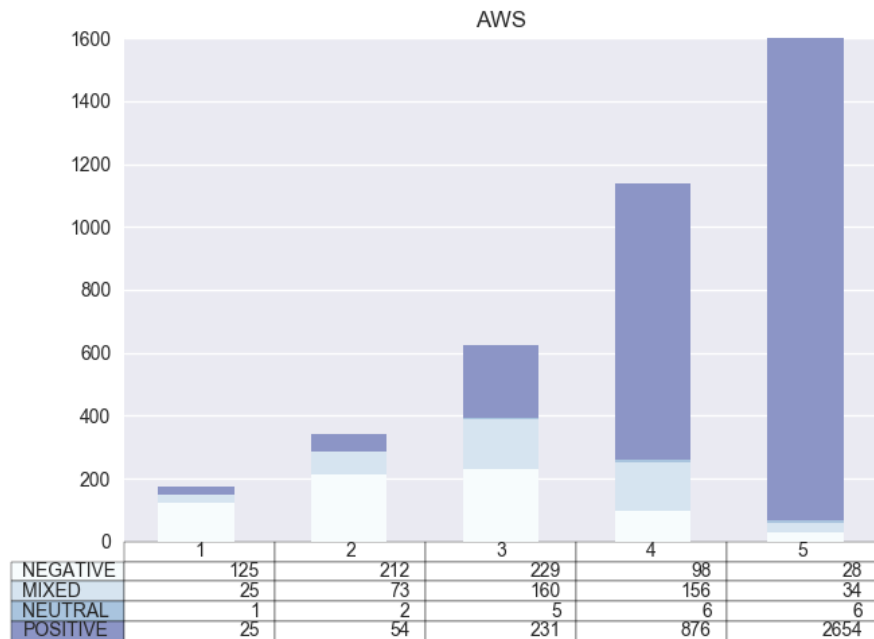


Abbildung 4.9: Sentiment Verteilung bei AWS

Mit Hilfe der in Unterabschnitt 4.3.3 festgelegten Kategorien wurden anschließend die richtigen und falschen Vorhersagen der Dienste gezählt und anhand dieser Werte die Accuracy der Modelle errechnet (s. Tabelle 4.6). Dabei erzielte Googles Dienst mit 95% Accuracy den höchsten Wert und kann demnach am besten beurteilen, ob ein Text ein positives oder negatives Sentiment aufweist. Darauf folgen AWS und IBM, welche mit 91,7% bzw. 89% dicht beieinander liegen. Das Modell von Azure hat mit 82,9% Accuracy am schlechtesten abgeschnitten. Wenn man dazu mitberücksichtigt, dass Azure zusätzlich bei neutralem/gemischtem Sentiment große Probleme bei der Beurteilung hat (vgl. Abbildung 4.4/Abbildung 4.7), schneidet Azure bei der Qualität der Sentimentanalyse insgesamt am schlechtesten ab.

	AWS	Azure	GCP	IBM
Richtig	3620	3273	3752	3513
Falsch	327	674	195	434
Accuracy	0,917	0,829	0,95	0,89

Tabelle 4.6: Ergebnisse der Kategorisierung nach positivem oder negativem Sentiment

Tabelle 4.7: Laufzeiten Sentimentanalyse

Anbieter	Laufzeit in s
IBM	1060
AZURE	707
GCP	1178
AWS	732

Bei der Laufzeit (Tabelle 4.7) liegt Azure jedoch vorne und konnte am schnellsten die Anfragen bearbeiten, dicht gefolgt von Amazon. Google und IBM benötigten deutlich mehr Zeit für die Analysen, bieten dafür aber auch wie erwähnt die beste Accuracy. Ein weiterer Grund für die längere Laufzeit der beiden Anbieter könnte sein, dass sie zunächst die Sprache des übergebenen Textes erkennen müssen. Diese Aufgabe fällt bei Azure und AWS weg, da dies explizit mit übergeben werden muss.

# 5 Resümee

In diesem Kapitel werden die Ergebnisse der Arbeit gebündelt. Anschließend folgt ein Ausblick über die weitere Entwicklung cloudbasierter Machine Learning Dienste und zuletzt eine reflexive Bewertung dieser Arbeit.

## 5.1 Zusammenfassung

Machine Learning Dienste sind heutzutage ein fester Bestandteil von zahlreichen Cloud-Angeboten. Von automatisierten Drag & Drop Lösungen bis hin zu vollwertigen Plattformen ist das Angebot sehr weitreichend. Aktuell lassen sich drei Segmente abgrenzen, die von allen vier untersuchten Anbietern abgedeckt werden.

Plattformbasierte Dienste stellen eine verwaltete Umgebung zu Verfügung, bei der man mit verschiedenen Frameworks eigene Modelle aufbauen und antrainieren kann. Die angebotenen Frameworks unterscheiden sich von Anbieter zu Anbieter, decken in ihrer Funktionalität jedoch dieselben Anwendungsfälle ab. Dadurch fällt ein Vergleich in dieser Dienste schwer. Eine Gegenüberstellung der Frameworks selbst, die die gleichen Funktionen anbieten, ist empfehlenswerter. Nichtsdestotrotz heben sich Amazon und Google deutlich bei den angebotenen SDKs hervor und unterstützen im Vergleich zu IBM und Azure mehrere Programmiersprachen. IBM und Azure bieten dagegen lediglich Unterstützung für Python.

Bei den automatisierten Diensten fällt eine eindeutige Gegenüberstellung ebenfalls schwer. Die Dienste unterscheiden sich sehr stark in der Benutzbarkeit und den Anwendungsgebieten. Azure bietet in dieser Kategorie jedoch die meisten Funktionen und unterstützt neben Regression und Klassifizierung zusätzliche Analyseverfahren. Google hatte sich zunächst mit einem Verweis auf ihren Plattformdienst aus diesem Bereich zurückgezogen. Seit neuestem werden jedoch Dienste zum automatischen antrainieren eigener Modelle

für die Bild- und Sprachanalyse angeboten. IBM und Amazon bieten Tools für Regression und Klassifizierung, die vollständig automatisiert und leicht zu nutzen sind.

Die vortrainierten ML APIs lassen sich am besten miteinander vergleichen, da neben dem Funktionsumfang die Qualität der dahinter liegenden Modelle untersucht werden kann. Dieser Dienstbereich lässt sich wiederum in drei Kategorien einteilen: Sprach- & Textanalyse, Bildanalyse, und andere Dienste. Letztere lassen sich schlecht miteinander vergleichen, da es sich zu einem um sehr spezialisierte und zum anderen Anbieter exklusive Dienste handelt. Daher fokussiert sich diese Arbeit auf die Dienste der Sprach- und Bildanalyse. Um diese genauer zu untersuchen wurden Tests für anbieterübergreifende Funktionen entwickelt, anhand derer Ergebnisse eine Evaluierung durchgeführt wurde. Für die Bildanalyse sind die ausgewählten Funktionen die Gesichts- und Geschlechtererkennung, für die Sprachanalyse die Sentimentbestimmung.

Für die Gesichts- und Geschlechtererkennung wurde ein Datensatz aus Bildern mit menschlichen Gesichtern und welchen ohne zusammengestellt. Die verschiedenen Dienste haben diese analysiert und wurden mithilfe klassifizierungsspezifischen Metriken bewertet. Bei der Gesichtserkennung liegt Google an erster Stelle und konnte als einziger Anbieter alle Bilder korrekt klassifizieren. Amazon und Azure liegen in der Accuracy gleich auf, jedoch weist das Modell von Azure eine bessere Precision auf; das von AWS einen besseren Recall. Daher kann je nach Anwendungsfall die Präferenz zwischen diesen beiden Anbietern unterschiedlich sein. IBM liegt bei allen drei Metriken an letzter Stelle. Bei der Geschlechtererkennung liegt Azure mit 100% Accuracy vorne, gefolgt von IBM mit 98,5% und AWS mit 98%. Der Dienst von Google konnte nicht bewertet werden, da diese Funktion nicht unterstützt wird. Hinsichtlich des Funktionsumfangs für die Bildanalyse führen AWS und Azure, welche zusätzliche Features für eine detailliertere Gesichtsanalyse anbieten.

Um die Sentiment Analyse zu bewerten wurde ein anonymisierter E-Commerce Datensatz verwendet. Die Dienste haben mittels der Kundenrezension das Sentiment bestimmt. Für die Auswertung wurden die Modelle als binäre Klassifikatoren betrachtet, welche eine Rezension entweder als positiv oder negativ einstufen. Die Bewertung der Dienste erfolgt mittels der Accuracy. Google konnte mit 95% Accuracy am besten das Sentiment der Rezensionen bestimmen. Jedoch wurden für einige Rezensionen keine Sentiment Werte zurückgegeben und das ohne eine Fehlermeldung oder einen Hinweis. Das Modell von AWS liefert eine Accuracy von 91,7%, gefolgt von IBM mit 89%. Azure schneidet bei der Sentiment Analyse am schlechtesten ab, das Modell bietet lediglich 82,9% Accuracy.



Letztendlich übertreffen Googles Dienste alle anderen Anbieter hinsichtlich der betrachteten Metriken. Amazon folgt testübergreifend an zweiter Stelle und bietet bei der Bildanalyse zusätzliche Funktionen gegenüber Google. Azure bietet insgesamt den größten Funktionsumfang bei den vortrainierten ML APIs und kann sich zusätzlich mit den automatisierten Diensten hervorheben. IBM scheint sich bei den vortrainierten Modellen mehr auf die Sprach- & Textanalyse zu konzentrieren und bietet dort die meisten Funktionen, kann aber bezüglich der unterstützten Sprachen nicht mit Google mithalten.

## 5.2 Ausblick

Man kann davon ausgehen, dass cloudbasierte ML Dienste auch in Zukunft an Funktionalität und Qualität dazugewinnen werden. Laut einer Studie der Infoholic Research LLP [2017] soll der MLaaS Markt im Prognosezeitraum 2017-2023 um 49% wachsen.

Ein potenzieller und viel versprechender Anwendungsbereich für MLaaS ist das **Internet of Things** (IoT). Dieses soll sich bis 2020 aus über 20 Mrd. Geräten (ausgenommen PCs, Tablets und Smartphones) zusammensetzen (Gartner, Inc. [2015]). Da die meisten Cloud Provider bereits spezielle Dienste für das IoT anbieten, welche sich leicht mit anderen Cloud-Diensten integrieren lassen, könnte so MLaaS eine Schlüsselrolle in diesem Bereich einnehmen.

Zukünftige Arbeiten könnten die Dienste anderer Kategorien genauer untersuchen und für diese Evaluierungsverfahren entwickeln oder die Dienste in einem anderen Fokus (z.B. wirtschaftlich) betrachten. Ein weiterer interessanter Punkt ist der Aufbau vortrainierter Modelle und in welchem Maß man die dahinter liegenden Lernverfahren untersuchen kann. Ein weiteres aktuelles Thema beschäftigt sich mit dem „Diebstahl“ von ML Modellen. Kesarwani et al. [2017] haben ein Verfahren aufgezeigt, mit denen ML Modelle durch gezielte API Anfragen extrahiert werden können. Die vortrainierten Modelle bilden die Grundlage für lukrative Cloud-Dienste; durch extrahierte Modelle könnten jedoch böswillige Nutzer die Dienste nachstellen oder zumindest API Anfragen und somit Kosten einsparen. Weiterführende Forschungen könnten diesbezüglich die Anbieter auf ihre Anfälligkeit gegenüber diesem Angriffsszenario untersuchen.

### 5.3 Reflexion

Die Arbeit sollte die aktuell verfügbaren cloudbasierten Machine Learning Dienste untersuchen und miteinander vergleichen. Die Anbietersauswahl im Vorhinein einzuschränken half dabei, die Übersicht bei den vielfältigen Dienstangeboten und API-Referenzen zu behalten. Wenn im Voraus bereits die zu evaluierenden Dienstkategorien festgestanden hätten, hätte man bei der Anbietersauswahl eventuell noch einen spezialisierten Anbieter für die Gesichts- oder Sentimentanalyse hinzunehmen können. So beurteilt die Arbeit lediglich Anbieter ohne spezifischen Fokus.

Den Schwerpunkt der Evaluierung auf die vortrainierten ML APIs zu setzen erwies sich als eine gute Entscheidung, da die Anbieter dort ähnliche Funktionen unterstützen und die dazugehörigen Tests interessant und von aktueller Bedeutung sind. Alternative wäre es auch aufschlussreich gewesen zu untersuchen, wie ein bestimmtes Framework auf den unterschiedlichen plattformbasierten Diensten abschneidet.

Python eignete sich hervorragend für die Entwicklung und Auswertung der Tests. Alle ausgewählten Anbieter bieten ein Python-SDK für die Anbindung und Nutzung ihrer Dienste an. Und auch die Visualisierung der Ergebnisse war dank der vorhandenen Python-Bibliotheken mühelos.

# Literaturverzeichnis

- Amazon Web Services, Inc. What Is Amazon Lex? - Amazon Lex, 2018a. URL <https://docs.aws.amazon.com/lex/latest/dg/what-is.html>. Zugriffsdatum: 2018-09-10.
- Amazon Web Services, Inc. What Is Amazon Comprehend? - Amazon Comprehend, 2018b. URL <https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html>. Zugriffsdatum: 2018-09-07.
- Amazon Web Services, Inc. What Is Amazon Polly? - Amazon Polly, 2018c. URL <https://docs.aws.amazon.com/polly/latest/dg/what-is.html>. Zugriffsdatum: 2018-09-07.
- Amazon Web Services, Inc. API Reference - Amazon Rekognition, 2018d. URL [https://docs.aws.amazon.com/rekognition/latest/dg/API\\_Reference.html](https://docs.aws.amazon.com/rekognition/latest/dg/API_Reference.html). Zugriffsdatum: 2018-09-13.
- Amazon Web Services, Inc. What Is Amazon Transcribe? - Amazon Transcribe, 2018e. URL <https://docs.aws.amazon.com/transcribe/latest/dg/what-is-transcribe.html>. Zugriffsdatum: 2018-09-07.
- Amazon Web Services, Inc. What Is Amazon Translate? - Amazon Translate, 2018f. URL <https://docs.aws.amazon.com/translate/latest/dg/what-is.html>. Zugriffsdatum: 2018-09-07.
- N. Brooks. Women's E-Commerce Clothing Reviews | Kaggle, 2018. URL <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>. Zugriffsdatum 2018-10-12.
- D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. pages 3642–3649, Providence, RI, June 2012. IEEE. doi: 10.1109/CVPR.2012.6248110. URL <http://ieeexplore.ieee.org/document/6248110/>.

- L. Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, Apr. 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.79. URL <http://ieeexplore.ieee.org/document/1597116/>. Zugriffsdatum: 2018-11-12.
- R. T. Fielding. *Architectural styles and the design of network-based software architectures*, volume 7. University of California, Irvine Irvine, USA, 2000.
- J. Fingas. Airports may use face recognition to screen US citizens (update: more info), May 2017. URL <https://www.engadget.com/2017/05/09/airport-face-recognition-for-us-citizens/>. Zugriffsdatum: 2018- 10-12.
- Foursquare Labs, INC. Versioning, 2018. URL <https://developer.foursquare.com/docs/api/configuration/versioning>. Zugriffsdatum: 2018-10-23.
- Gartner, Inc. Gartner Says 6,4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015. Nov. 2015. URL <https://www.gartner.com/newsroom/id/3165317>. Zugriffsdatum: 2018-11-19.
- Gartner, Inc. Gartner Forecasts Worldwide Public Cloud Revenue to Grow 17.3 Percent in 2019, 2018. URL <https://www.gartner.com/en/newsroom/press-releases/2018-09-12-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2019>. Zugriffsdatum: 2019-01-07.
- Google LLC. Global Locations - Regions & Zones, 2018a. URL <https://cloud.google.com/about/locations/>. Zugriffsdatum: 2018-10-12.
- Google LLC. Cloud AutoML - Custom Machine Learning Models | AutoML, 2018b. URL <https://cloud.google.com/automl/>. Zugriffsdatum: 2018-09-15.
- Google LLC. Human labeling | Cloud AutoML Vision, 2018c. URL <https://cloud.google.com/vision/automl/docs/human-labeling>. Zugriffsdatum: 2018-09-15.
- Google LLC. Cloud Natural Language | Cloud Natural Language, 2018d. URL <https://cloud.google.com/natural-language/overview/docs/>. Zugriffsdatum: 2018-09-07.
- Google LLC. Cloud Speech-to-Text Documentation | Cloud Speech-to-Text API, 2018e. URL <https://cloud.google.com/speech-to-text/docs/>. Zugriffsdatum: 2018-09-07.

- Google LLC. Cloud Text-to-Speech API documentation | Cloud Text-to-Speech API, 2018f. URL <https://cloud.google.com/text-to-speech/docs/>. Zugriffsdatum: 2018-09-07.
- Google LLC. Cloud Talent Solution - Job Matching APIs | Talent Solution, 2018g. URL <https://cloud.google.com/solutions/talent-solution/>. Zugriffsdatum: 2018-09-10.
- Google LLC. Cloud Translation | Cloud Translation, 2018h. URL <https://cloud.google.com/translate/overview/docs/>. Zugriffsdatum: 2018-09-07.
- Google LLC. Cloud Video Intelligence API | Cloud Video Intelligence API, 2018i. URL <https://cloud.google.com/video-intelligence/docs/reference/rest/>. Zugriffsdatum: 2018-09-13.
- Google LLC. Method: images.annotate | Cloud Vision API, 2018j. URL <https://cloud.google.com/vision/docs/reference/rest/v1/images/annotate>. Zugriffsdatum: 2018-09-13.
- M. Hossin and M. N. Sulaiman. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2):01–11, Mar. 2015. ISSN 2231007X, 22309608. doi: 10.5121/ijdkp.2015.5201. URL <http://www.airconline.com/ijdkp/V5N2/5215ijdkp01.pdf>.
- G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, Oct. 2007.
- IBM Corp. IBM Cloud Docs, 2018a. URL <https://console.bluemix.net/docs/services/assistant/index.html#about>. Zugriffsdatum: 2018-09-10.
- IBM Corp. IBM Cloud Docs, 2018b. URL <https://console.bluemix.net/docs/services/natural-language-understanding/index.html#about>. Zugriffsdatum: 2018-09-07.
- IBM Corp. IBM Cloud Docs, 2018c. URL <https://console.bluemix.net/docs/services/personality-insights/index.html#about>. Zugriffsdatum: 2018-09-07.
- IBM Corp. IBM Cloud Docs, 2018d. URL <https://console.bluemix.net/docs/services/speech-to-text/index.html#about>. Zugriffsdatum: 2018-09-07.

- IBM Corp. IBM Cloud Docs, 2018e. URL <https://console.bluemix.net/docs/services/text-to-speech/index.html#about>. Zugriffdatum: 2018-09-07.
- IBM Corp. IBM Cloud Docs, 2018f. URL <https://console.bluemix.net/docs/services/tone-analyzer/index.html#about>. Zugriffdatum: 2018-09-07.
- IBM Corp. IBM Cloud Docs, 2018g. URL <https://console.bluemix.net/docs/services/language-translator/index.html#about>. Zugriffdatum: 2018-09-07.
- IBM Corp. Visual Recognition - IBM Cloud API Docs, 2018h. URL <https://cloud.ibm.com/apidocs/visual-recognition>. Zugriffdatum: 2018-09-13.
- Infoholic Research LLP. Machine Learning as a Service Market - Global Drivers, Restraints, Opportunities, Trends, and Forecasts up to 2023. Technical Report 4418420, Sept. 2017. URL [https://www.researchandmarkets.com/research/czmprd/machine\\_learning](https://www.researchandmarkets.com/research/czmprd/machine_learning). Zugriffdatum: 2018-11-16.
- P. Jawandhiya. Hardware Design for Machine Learning. *International Journal of Artificial Intelligence & Applications*, 9(1):63–84, Jan. 2018. ISSN 09762191. doi: 10.5121/ijaia.2018.9105. URL <http://airconline.com/ijaia/V9N1/9118ijaia05.pdf>.
- M. Kesarwani, B. Mukhoty, V. Arya, and S. Mehta. Model Extraction Warning in MLaaS Paradigm. *arXiv:1711.07221 [cs]*, Nov. 2017. URL <http://arxiv.org/abs/1711.07221>. arXiv: 1711.07221.
- S. Kreml. CCC: Bundespolizei hat Bericht zur Gesichtserkennung absichtlich geschönt, Oct. 2018. URL <https://www.heise.de/newsticker/meldung/CCC-Bundespolizei-hat-Bericht-zur-Gesichtserkennung-absichtlich-geschoent-4191216.html>. Zugriffdatum: 2018-10-29.
- B. Marr. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read, May 2018. URL <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>. Zugriffdatum: 2018-11-06.
- M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural*

- Networks: The Official Journal of the International Neural Network Society*, 16(5-6): 555–559, July 2003. ISSN 0893-6080. doi: 10.1016/S0893-6080(03)00115-1.
- P. Mell and T. Grance. The NIST Definition of Cloud Computing. Technical Report NIST Special Publication (SP) 800-145, Sept. 2011. URL <https://csrc.nist.gov/publications/detail/sp/800-145/final>.
- Microsoft Corp. Text Analytics API Documentation - Tutorials, API Reference - Azure Cognitive Services, 2018a. URL <https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/>. Zugriffsdatum: 2018-09-08.
- Microsoft Corp. Bing Spell Check API Documentation - Tutorials, API Reference - Azure Cognitive Services, 2018b. URL <https://docs.microsoft.com/en-us/azure/cognitive-services/bing-spell-check/>. Zugriffsdatum: 2018-09-08.
- Microsoft Corp. Azure Bot Service - Bot Service, 2018c. URL <https://docs.microsoft.com/en-us/azure/bot-service/>. Zugriffsdatum: 2018-09-10.
- Microsoft Corp. What are Azure Cognitive Services?, 2018d. URL <https://docs.microsoft.com/en-us/azure/cognitive-services/welcome>. Zugriffsdatum: 2018-09-08.
- Microsoft Corp. Face API - V1.0, 2018e. URL <https://westus.dev.cognitive.microsoft.com/docs/services/563879b61984550e40cbb8d/operations/563879b61984550f30395236>. Zugriffsdatum: 2018-09-13.
- Microsoft Corp. What is the Linguistic Analysis API, 2018f. URL <https://docs.microsoft.com/en-us/azure/cognitive-services/linguisticanalysisapi/home>. Zugriffsdatum: 2018-09-08.
- Microsoft Corp. Azure Search Documentation – Tutorials, quickstarts, API references, 2018g. URL <https://docs.microsoft.com/en-us/azure/search/>. Zugriffsdatum: 2018-09-10.
- Microsoft Corp. What is the Speaker Recognition API? - Azure Cognitive Services, 2018h. URL <https://docs.microsoft.com/en-us/azure/cognitive-services/speaker-recognition/home>. Zugriffsdatum: 2018-09-08.
- Microsoft Corp. Azure Machine Learning Studio Algorithm and Module Reference, 2018i. URL <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/>. Zugriffsdatum: 2018-09-15.

- Microsoft Corp. What is the Translator Speech service? - Azure Cognitive Services, 2018j. URL <https://docs.microsoft.com/en-us/azure/cognitive-services/translator-speech/overview>. Zugriffdatum: 2018-09-08.
- Microsoft Corp. Translator Text Documentation - Quickstarts, Tutorials, API Reference - Azure Cognitive Services, 2018k. URL <https://docs.microsoft.com/en-us/azure/cognitive-services/translator/>. Zugriffdatum: 2018-09-08.
- Project Jupyter. The Jupyter Notebook, 2019. URL <https://www.jupyter.org>. Zugriffdatum: 2019-01-08.
- J. Redmon. How computers learn to recognize objects instantly [Video file], Apr. 2017. URL [https://www.ted.com/talks/joseph\\_redmon\\_how\\_a\\_computer\\_learns\\_to\\_recognize\\_objects\\_instantly](https://www.ted.com/talks/joseph_redmon_how_a_computer_learns_to_recognize_objects_instantly). Zugriffdatum: 2018-09-12.
- RightScale, Inc. RightScale 2018 State of the Cloud Report, 2018. URL <https://www.rightscale.com/lp/state-of-the-cloud>. Zugriffdatum: 2019-01-07.
- R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, Feb. 1978. ISSN 00010782. doi: 10.1145/359340.359342. URL <http://portal.acm.org/citation.cfm?doid=359340.359342>.
- S. Schuermans, A. Sobolevski, C. Voskoglou, and M. Wilcox. State of the Developer Nation Q1 2017. Technical report, VisionMobile, 2017. URL <https://www.visionmobile.com/free-resources/state-developer-nation-q1-2017>.
- A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499 [cs]*, Sept. 2016. URL <http://arxiv.org/abs/1609.03499>.
- T. Young, D. Hazarika, S. Poria, and E. Cambria. Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. *IEEE Computational Intelligence Magazine*, 13(3):55–75, Aug. 2018. ISSN 1556-6048. doi: 10.1109/MCI.2018.2840738. URL <https://ieeexplore.ieee.org/document/8416973/>.



## Erklärung zur selbstständigen Bearbeitung einer Abschlussarbeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg

11.01.2019



Ort

Datum

Unterschrift im Original