

Interdisziplinäres Projekt

vorgelegt von

Halil Kocak

Bachelor-Studiengang Informatik

Fakultät Vermessung, Informatik und Mathematik

Wintersemester 2021/22

**Ausarbeitung des Modells für Bildklassifizierung
nach CRISP-DM**

Inhaltsverzeichnis

Abbildungsverzeichnis	3
Tabellenverzeichnis	4
1 Einleitung	5
2 Maschinelles Lernen	6
2.1 Überwachtes Lernen	7
2.2 Künstliche Neuronale Netze	7
2.2.1 Neuronales Netz	7
2.2.2 Feed-Forward Propagierung	8
2.2.3 Eingabedaten	10
2.2.4 Lernverfahren	11
2.3 Deep Learning	12
2.3.1 Convolutional Neural Network	13
2.4 Evaluationstechniken	14
3 CRISP-DM Modell	15
4 Experimente	18
4.1 Business Understanding	18
4.2 Data Understanding	18
4.3 Data Preparation	21
4.3.1 Test- und Trainingsdaten	21
4.3.2 Skalierung der Merkmale	21
4.3.3 Datentransformation	22
4.4 Modellierung	22
4.4.1 Experiment 1	22
4.4.2 Experiment 2	24
Literaturverzeichnis	25

Abbildungsverzeichnis

2.1	Maschinelles Lernen vs. Klassische Programmierung [Cho18](K.1.1.2	6
2.2	Modell eines Neurons [Abbe]	8
2.3	Schichten der Künstliche Neuronale Netze [Abbf]	8
2.4	Feed-Forward Netz [Hal22]	10
2.5	Künstliches Neuronales Netz vs. Deep Learning [Abbd]	13
2.6	Aufbau eines CNNs [Abba]	14
3.1	Ablauf CRIPS-DM Modell [Abbb]	15
4.1	Der gesamte Datensatz: FASHION-MNIST [Abbc]	19
4.2	Verteilung Trainingsdatensatz [Abbh]	20
4.3	Verteilung Testdatensatz [Abbg]	21
4.4	Experiment 1: Modellaufbau [HK221] (Experiment 1)	23
4.5	Experiment 1: Bewertung Trainingsphase [HK221](Experiment 1)	24

Tabellenverzeichnis

4.1	Labels des Datensatzes	20
-----	----------------------------------	----

1 Einleitung

Es war schon immer die Frage ob Maschinen jemals selbst, anhand vorhandenen Informationen, Entscheidungen treffen oder wie ein Mensch Regeln aus bestimmten Prozessen extrahieren können [Cho18](K.1.1.1). Diese Fragestellung ist vor allem in der heutigen Zeit, in der das Thema Digitalisierung voran getrieben wird, von großer Bedeutung. Auch setzen Unternehmen immer mehr auf Techniken aus dem Bereich des Maschinellen Lernens (ML) und verwenden diese um ihre Prozesse zu digitalisieren und somit zu optimieren. Es existieren verschiedene ML-Techniken, wie zum Beispiel die Klassifikation, Regression oder Deep Learning. Letzteres steht im Mittelpunkt vieler Anwendungsszenarien. Anwendungsfelder für Deep Learning sind vor allem, Bild- und Videoanalyse, sowie die Sprachverarbeitung. Diese werden in verschiedenen Branchen, wie zum Beispiel der Medizin, der Automobilindustrie oder im Kundendienst eingesetzt [FG18]. Auch im Bereich Online-Retail finden ML-Techniken einen relevanten Einsatz. Im Rahmen dieser Bachelor-Arbeit liegt der Fokus vor allem in diesem Bereich und wie die Bildanalyse hier unterstützen kann. Diese Problemstellung wird in einem späteren Abschnitt der Arbeit näher definiert.

Diese Bachelor-Arbeit beinhaltet die Erkennung von Objekten auf Bildern durch den Einsatz von Künstlicher Intelligenz. Dabei wird die Disziplin Deep Learning unter Einsatz von Convolutional Neural Network verwendet. Um diese zu verstehen werden allgemeine Informationen zur Künstlichen Intelligenz und dessen Bestandteile und Algorithmen benötigt. Zu Beginn ist es wichtig zu wissen, welche Probleme die Künstliche Intelligenz lösen kann und welche Bestandteile für diese Arbeit benötigt werden. In diesem Zusammenhang wird zum Beispiel das Verfahren Convolutional Neural Network behandelt, welches zu den Deep Learning Methoden gehört und eine Teilmenge der Maschinellen Lernalgorithmen bildet. Außerdem wird im Rahmen dieser Arbeit das CRIPS-DM Modell vorgestellt und verwendet, welches als Standard für Data Mining Modelle definiert wurde.

2 Maschinelles Lernen

Maschinelles Lernen stellt ein Teilgebiet der Künstlichen Intelligenz dar. Dabei werden durch den Einsatz von Maschinen ein Verhalten aus den Daten extrahiert, wodurch die Maschine selbst Lernen kann. Dies geschieht unter Verwendung verschiedener Techniken und Algorithmen. Während der Lernphase versucht die Maschine, Regelmäßigkeiten aus den Daten zu erkennen und diese zu verallgemeinern. Nachdem die Lernphase beendet ist, kann die Maschine eine Vorhersage aus den aktuellen Daten treffen [Fro21](S.2-3).

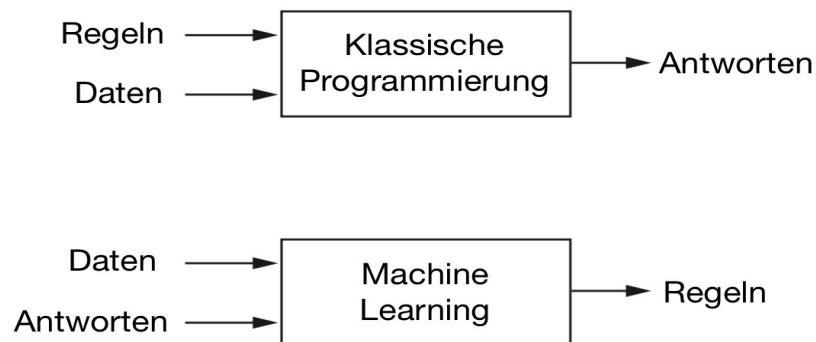


Abbildung 2.1: Maschinelles Lernen vs. Klassische Programmierung [Cho18](K.1.1.2)

In der Abbildung 2.1, ist der Unterschied zwischen der Klassischen Programmierung und den Maschinellen Lernalgorithmen zu sehen. Bei der Klassischen Programmierung werden Regeln definiert und implementiert. Anhand dieser Regeln werden die Daten verarbeitet, die dann zu einer Antwort führen. Anders ist es aber bei den Maschinellen Lernalgorithmen. Hier werden Daten und dessen Antworten als Eingabedaten betrachtet, wodurch die Maschine für sich die Regeln definiert [Cho18](Kapitel 1). Dazu haben alle Maschinelle Lernverfahren zwei Phasen. Die erste Phase ist die Lernphase, in der es lernt, aus den Eingabedaten ein Muster zu extrahieren. Nachdem diese Phase beendet ist, wird die Anwendungsphase gestartet. Diese Phase wird sowohl für die Evaluation des Künstliche Intelligenz als auch für die Anwendung genutzt. Es ist zu beachten das die Künstliche Intelligenz auch in der Anwendungsphase lernen kann [Cho18](Kapitel 2.1).

Um Entscheidungen zu treffen, werden Eingabedaten benötigt, die der Maschine die nötigen Informationen liefern. Die Eingabedaten beinhalten auch die Lösung des Problems und somit auch den Wert, der von der Maschine vorhergesagt werden soll. Damit die Maschine nach Effizienz bewertet werden kann, ist eine Messung der Algorithmen erforderlich [Cho18](Kapitel 1.1.3).

Für die Lösungen der Probleme werden verschiedentliche Maschinelle Lernalgorithmen verwendet. Die Entscheidung welches benutzt werden soll hängt von den Eingabedaten und dem Ziel ab. Im Rahmen dieser Arbeit liegt der Fokus auf dem Convolutional Neural Network Verfahren, welches zur Kategorie des überwachten Lernens gehört [Lub19].

2.1 Überwachtes Lernen

Das überwachte Lernen gehört zu den Lernverfahren des maschinellen Lernens und umfasst die Klassifikation und Regression. Bei einem überwachten Lernen werden Eingabedaten, die Beispiele beinhalten, einem bekannten Zielwert zugeordnet. Dadurch sind die Zielwerte bekannt und somit können die Fehler, bei einer falschen Klassifikation, während dem Lernprozess entdeckt werden [Cho18](K.4.1.2).

2.2 Künstliche Neuronale Netze

Im folgenden Abschnitt werden die Bestandteile von künstlichen Neuronalen Netzen definiert und der Aufbau beschrieben.

2.2.1 Neuronales Netz

Auf der linken Seite der Abbildung 2.2, befinden sich die Eingabedaten die als $x_0, x_1, \dots, x_n \in \vec{X}$ definiert sind, wobei \vec{X} den Vektor \mathbb{R}^n darstellt. Jede einzelnen Eingabe x_0, x_1, \dots, x_n wird mit seinem Gewicht $w_{n,j}$ multipliziert, wobei $j \in [0, AnzahlNeuronen]$ gilt. Nachdem multiplizieren mit den Gewichten werden alle berechneten Werte an die Übertragungsfunktion übergeben. Dieser Ablauf ist dabei an eine echte Nervenzelle angelehnt, indem Neuronen aus anderen Nervenzellen, dieses Neuron anregen und auf diese übertragen können. Die Übertragungsfunktion summiert diese Werte zusammen und leitet sie dann an die Aktivierungsfunktion weiter. Die Aktivierungsfunktion beinhaltet genauso ein Schwellenpotenzial mit der wir auch Schwellenwerte prüfen können. Eine einfache Aktivierungsfunktion wäre hierbei das Alles-oder-nichts-Prinzip. Mit diesem Prinzip ist gemeint, dass bei

Überschreitungen des Schwellwertes das Neuron aktiviert und unterhalb des Schwellwertes inaktiv gestellt wird [Fro21](S.169).

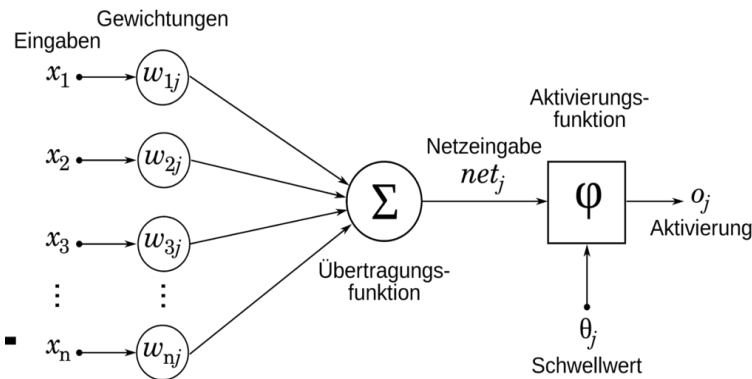


Abbildung 2.2: Modell eines Neurons [Abbe]

Zunächst ist wichtig wie die Vernetzungen und Verknüpfungen der Neuronen aussehen und funktionieren. In Abbildung 2.3 wird deutlich, dass ein einfaches Künstliches Neuronales Netz aus drei Schichten besteht. Die erste Schicht ist dabei die Eingabeschicht (engl. Inputlayer), die die Eingabe der Daten verarbeitet und wiederum mit jedem Neuron aus der verborgenen Schicht (engl.= Hiddenlayer) vernetzt ist. Durch die Ausgabeschicht werden Regeln, die aus der verborgenen Schicht extrahiert werden, bereitgestellt. Diese Neuronen werden zur Evaluation des Modells benutzt [Fro21](S.169).

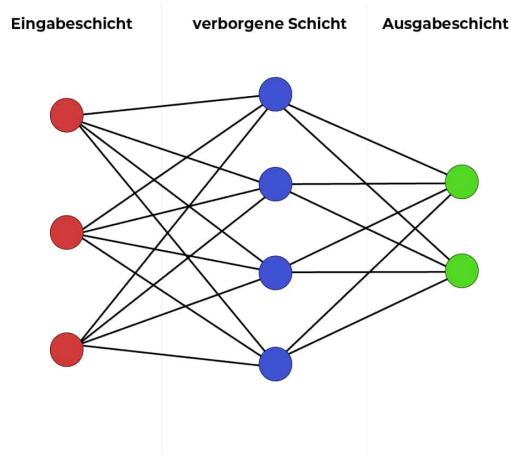


Abbildung 2.3: Schichten der Künstliche Neuronale Netze [Abbf]

2.2.2 Feed-Forward Propagierung

Das neuronale Netz durchläuft bei jedem Dateieintrag einmal alle Schichten. Verwendet wird dafür, dass im Abbild 2.2 vorgestellte Modell.

Mathematische Beschreibung eines Feed-Forward Netzes

In den vorherigen Abschnitten wurden die Bestandteile und deren Funktionen näher erklärt. An dieser Stelle wird das einfache neuronale Netz beschrieben. Dafür wird zuerst die Arbeitsweise zwischen den Schichten definiert:

$$f : \mathbb{R}^{D_0} \rightarrow \mathbb{R}^{D_l}, f(\vec{X}) = (f_l \circ f_{l-1} \circ \dots \circ f_1)(u),$$

wobei f_l die Schichten sind und für jedes $L \in \{0, 1, 2, \dots, l\}$ die Funktion,

$$f : \mathbb{R}^{D_{L-1}} \rightarrow \mathbb{R}^{D_L}, f_L(\vec{X}) = \phi(w^L * \vec{X} + b^L), \text{ beschreibt.}$$

Wobei \vec{X} der Vektor für die Eingabedaten (engl. Inputdata), $w^L \in \mathbb{R}^{(D_L \times D_{L-1})}$ die Gewichte, welche die Verbindungen (Abbild 2.3) von einem Neuron zu anderen Neuronen gewichten und b^L der Verzerrungsvektor (engl. Bias), der Schicht L ist. ϕ ist die Aktivierungsfunktion der Schicht [Pla21] (S.232 - 234).

Lern- und Testphase

Während der Lernphase wird das Model mit einer Teilmenge von den gesamten Daten, gefüttert. Durch dies wird das Model trainiert, daher nennt man diese Phase auch Trainingsphase. Damit keine Informationen aus der gesamten Datenmenge durchsickern, werden nur die Trainingsdaten in das Model gefüttert. Damit das Model während der Trainingsphase bewertet werden kann, benötigt es eine Menge an Daten aus dem gesamten Datensatz, die als Validierungsmenge an das Model übertragen wird. Somit stellt sich fest, dass bei der Bewertung eines Modells, damit anfängt die gesamte Daten in Trainings-, Validierungs- und Testdaten zu verteilen [Cho18] (K.4.2.1).

Die Gewichte werden zunächst zufällig vergeben und mit dem Lernverfahren, das während der Rückpropagierung stattfindet, optimiert. Damit das Neuronale Netz Abweichungen erkennen kann, werden Bias verwendet, die diese Abweichung darstellen. Bevor mit der Trainingsphase begonnen wird, sollten folgende Elemente festgelegt werden. Zunächst ist die Verlustfunktion zu definieren, welche die Leistung der Trainingsdaten beschreibt. Durch die Verlustfunktion kann das Neuronale Netz die Gewichte optimieren. Als nächstes sollte die Menge an Trainingsdaten und Testdaten festgelegt werden [Cho18] (K.2.1 Listing 2.2). An dieser Stelle werden beispielhaft die Schritte der Feed-Forward Propagation in Abbild 2.4 betrachtet. Zunächst wird die Eingabeschicht mit den Eingabedaten gefüttert, wobei diese normalisierte Werte beinhalten sollten. Die zufällig generierten Gewichte stellen die

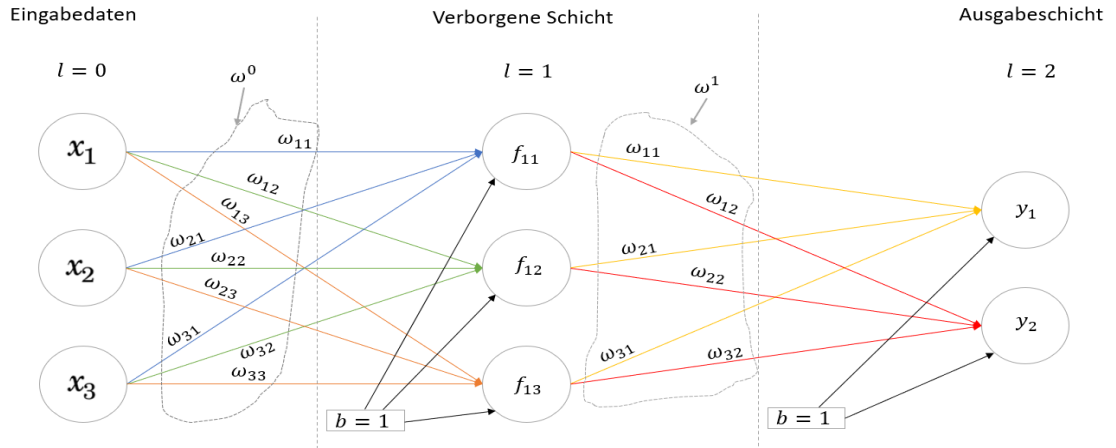


Abbildung 2.4: Feed-Forward Netz [Hal22]

Gewichtungen für die Neuronen der nächsten Schicht dar. Nun wird das, in Abbild 2.4 definierte, Model für jedes Neuron aus der nächsten Schicht angewendet. Somit hat das erste Neuron aus der Schicht $l=1$ folgende Berechnung.

$$f_{11} = \phi[(x_1 * w_{11}) + (x_2 * w_{21}) + (x_3 * w_{31})] + b^1$$

allgemein lässt sich die Feed-Forward Propagierung wie folgt beschreiben:

$$f = \sum_{n=0}^n [\phi(x_n * w_{nj} + b^j)]$$

, wobei n die Anzahl an Neuronen im vorherigen Schicht sind und j die Anzahl an Neuronen im aktuellen Schicht. Somit kann das Neuronale Netz mit den bestmöglichen Gewichten Prognosen anhand Eingabedaten, liefern. Die Optimierung der Gewichte wird im Abschnitt 2.2.4 Lernverfahren näher beschrieben.

2.2.3 Eingabedaten

Die Eingabedaten x_1, \dots, x_n stellen die Merkmale (oder Attribute) dar, welches die Eigenschaften beinhalten, aus denen das Neuronale Netz lernen kann. Zu beachten sind dabei die Datentypen der Eingabedaten und die Normierung dieser Daten. Dabei gibt es unterschiedliche Datentypen. Im Rahmen dieser Arbeit ist der Fokus auf Vektor- und Bilddaten gerichtet [Cho18] (K.2.2.8).

Vektordaten

Dieser Datentyp wird mathematisch als $\vec{x} \in \mathbb{R}^2$ definiert und beinhaltet in der ersten Achse, die Datenwerte, wie zum Beispiel x_1 . Die zweite Achse ist die Merkmalsachse, welche die Merkmale darstellt die zur Verfügung stehen [Cho18] (K.2.2.9).

Bilddaten

Bei den Bilddaten handelt es sich generell um drei dimensionale Matrizen, jedoch unterscheiden wir die Graustufenbilder, für diese zwei Dimensionale Matrizen ausreichen. Hier bekommt jeweils eine Achse die Höhe, die zweite Achse die Breite und die letzte Achse die Farbtiefe des Bildes. Da die Graustufenbilder nur einen Farbkanal beinhalten, ist die zwei dimensionale Matrix ausreichend [Cho18] (K.2.2.11).

2.2.4 Lernverfahren

Bisher wurde der Durchlauf eines Neuronalen Netzes definiert, jedoch nicht wie es aus den Daten lernen kann. Damit das Neuronale Netz lernen kann, muss er auch Fehler erkennen können. Abhängig von der Verlustfunktion ist auch die Aktivierungsfunktion, die für die Ausgabeschicht ausgewählt wurde. Je nach Größe oder Art der Ziel-Klassen wird eine Aktivierungsfunktion ausgewählt. Ein Beispiel dazu ist die Mehrfachklassifizierung, bei der die Aktivierungsfunktion SSoftmax ausgewählt wird. Die Softmax-funktion auch als Softmax-Regression bekannt, stellt anhand der Eingabewerte eine Wahrscheinlichkeitsverteilung zu Verfügung, deren Summe 1 ergibt. Somit werden den Ziel-Klassen Werte vergeben die zwischen 0 und 1 sind [Mah18]. Durch die Vereinheitlichung der Neuronen mit der Aktivierungsfunktion, können Fehlerwerte kalkuliert werden. Dies geschieht folgendermaßen. Es wird angenommen, dass die Ausgabeschicht den Wert $x_0 = 0.75$ beinhaltet, wobei der echte erwartete Wert $y = 1$ ist. Durch die Anwendung eines der Verlustfunktionen, stellt das Neuronale Netz fest, wie groß die Abweichung des Ergebnisses von dem tatsächlichen Ergebnis ist. Hierzu werden unterschiedliche Verfahren genutzt [Mat21](S.124-126). Eines dieser Verfahren ist die mittlere quadratische Abweichung (engl. mean squared error, MSE).

$$MSE = \frac{1}{n} \left(\sum_{i=1}^n \right) (Y_i - \hat{Y}_i)^2$$

, wobei n Anzahl der Datenpunkte, Y_n das erwartete Ergebnis und \hat{Y}_i die prognostizierten Werte sind.

In diesem Beispiel wird es wie folgt berechnet.

$$\begin{aligned} MSE &= \frac{1}{1} \sum_{i=1}^1 (Y_i - \hat{Y}_i)^2 \\ MSE &= 1 * (1 - 0.75)^2 = 0.0625 \end{aligned}$$

Somit kennt das Neuronale Netz die Abweichung der Prognose, womit es den Lernprozess starten kann. Es stellt sich die Frage wie das Neuronale Netz sich optimieren bzw. verbessern lässt. [Pla21] (S.194). Es ist bekannt dass ein Neuronales

Netz mehrere Ausgaben liefern kann und für jeder dieser Ausgaben ist es möglich eine Verlustfunktion zuzuweisen. [Cho18].

Verlustfunktion

Die Verlustfunktion (engl. loss-, costfunktion) wird in der Optimierung während der Trainingsphase genutzt. Hierzu gibt es Verschiedentliche Funktionen wie zum Beispiel der oben genannte mittlere quadratische Abweichung. Ein anderes Beispiel, ist die Cross-Entropy-Loss Funktion, welches für die Optimierung von klassifikatorische modele geeignet ist. Die Cross-Entropy-Loss Funktion ist ähnlich wie die Softmax Funktion aufgebaut. Sie misst den Abstand zwischen den Prognostizierten Wert und den Wahrheitswert, um so nah wie möglich in die richtige Richtung des Ergebnisses zu bewegen. Während der Trainingsphase werden die Gewichte iterativ und entsprechend mit dem Ziel, den Cross-Entropy-Loss zu minimieren, angepasst [Mah18]. Im Rahmen dieser Arbeit wird die Cross-Entropy-Loss Funktion angewendet, da es sich um eine mehrfachklassifizierung handelt.

Rückpropagierung

Die Anpassung der entsprechende Gewichte wird durch den Rückpropagierung ermöglicht. Damit die Gewichte angepasst werden können, ist es nötig das das Model Fehlerwert ein Minimum annimmt. Es ist bekannt das mit dem Gradienten der Funktion das Minimum erreichbar ist, welches die partielle Ableitung aller Vektoren darstellt. Dieser Vektor wird auch $\delta(f)$ genannt. Sie zeigt an jedem Punkt der Funktion, die Richtung der steilsten anstieg (bzw. maximum), wobei $-\delta(f)$ die Richtung der steilsten abstieg (bzw. minimum) der Funktion zeigt [uSV21] (S.337). Allgemein wird der Gradientenverfahren folgendermaßen definiert:

$$x^{i+1} = x^i - \eta * \delta(f(x^i))$$

, wobei x^0 der Startpunkt ist und η die Schrittweite. Diese Verfahren lässt sich wie folgt bei den Gewichten integrieren.

$$w_{neu} = w_{alt} - \eta * \delta(E(W_{alt}))$$

2.3 Deep Learning

Bei Deep Learning handelt es sich um eine Teilmenge der Maschinellen Lernalgorithmen, mit der komplexere Aufgaben gelöst und automatisiert werden können. Der Unterschied zu einem Künstlichen Neuronalen Netz ist, dass Deep Learning mehrere

verborgene Schichten beinhaltet (siehe Abbildung 2.5). Hierzu ist die Aktivierungsfunktion als die Sigmoid-Funktion vorgesehen, die wie folgt definiert ist:

$$\text{sig}(t) = \frac{1}{(1+e^{-t})}$$

,wobei $\text{sig}(t) \in [0, 1]$ und $t \in \mathbb{R}$ gilt.

Dies bedeutet, dass die Neuronen immer eine Zahl übertragen bekommen, die zwischen 0 und 1 liegt. Somit können durch Implementierungen Schwellenwerte definiert und dementsprechend Neuronen aktiviert oder deaktiviert werden [Fro21](S.178).

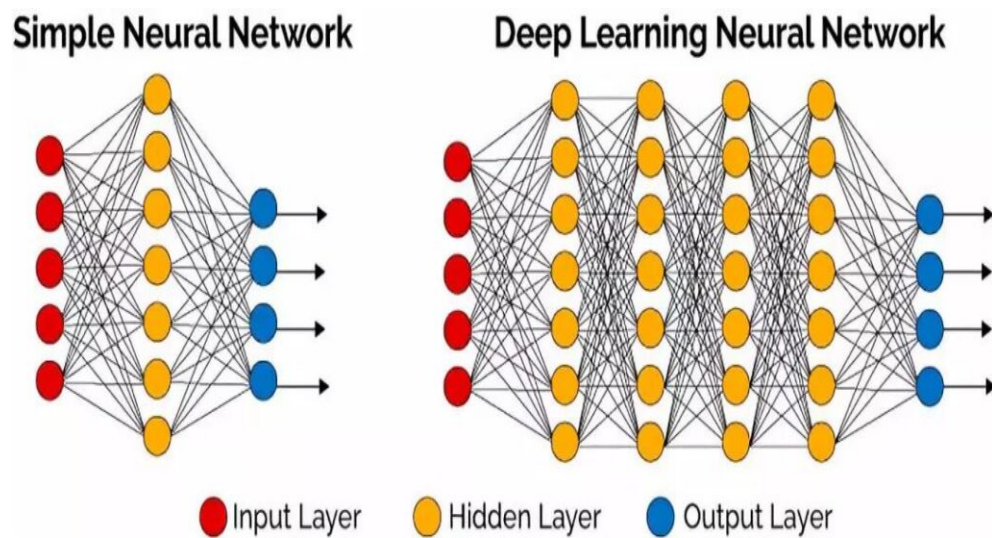


Abbildung 2.5: Künstliches Neuronales Netz vs. Deep Learning [Abbd]

2.3.1 Convolutional Neural Network

Bei einem CNN, im Deutschen "gefaltetes Neuronales Netzwerk", handelt es sich von einer Sonderform des künstlichen neuronalen Netzes, das speziell für Bild- oder Audibearbeitung entwickelt wurde. Diese Form ist teils an den biologischen Ablauf angelehnt.[Lub19] Der Aufbau eines CNNs ist in Abbildung 2.6 dargestellt.

In Abbildung 2.6 sind mehrere Schichten (Layer) zu sehen. Dabei ist die Eingabeschicht (Inputlayer) und die Ausgabeschicht (Outputlayer) bekannt. Wie im vorherigen Abschnitt auch definiert, hat ein CNN mehrere verborgene Schichten, darunter den **Convolutional Layer**, der einzelne Merkmale aus den Eingabedaten ortsunabhängig erkennt und daraufhin diese extrahiert. Diese Merkmale könnten bei einer Bildverarbeitung den Kanten, Linienformen und Farbtupfer entsprechen, die eine Struktur darstellen. Diese Merkmale werden in der **Pooling-Schicht**, auch

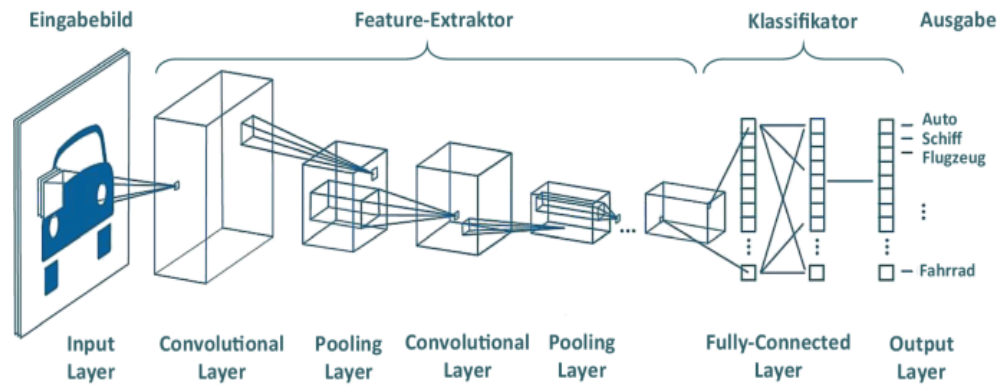


Abbildung 2.6: Aufbau eines CNNs [Abba]

als Subsampling bekannt, bearbeitet, indem die Auflösung der Merkmale reduziert und verdichtet wird. Somit werden überflüssige Informationen verworfen, wodurch die Datenmenge reduziert wird. Dadurch wird die Berechnungsgeschwindigkeit erhöht. Diese zwei Schichten können so oft wie möglich verwendet werden, da sie nur Merkmale aus den Bildern extrahieren und optimierend zu Verfügung stehen. Die eigentliche Klassifikation entsteht durch den **Fully-Connected Layer**, im Deutschen "vollständig verknüpfte Schicht", der sich der wiederholten Abfolge der Convolutional- und Pooling-Schichten anschließt, indem alle Merkmale der vorgelagerten Schichten mit jedem Ausgabemerkmale verknüpft sind. Somit werden mögliche Kombinationen aus den Strukturen entworfen, die komplexere Strukturen identifizieren lassen. Wichtig ist hierbei, dass die Anzahl an Neuronen in der vollständig vernetzten Schicht abhängig von den Klassen oder Objekten sind [Lub19].

2.4 Evaluationstechniken

3 CRISP-DM Modell

Das CRISP-DM Modell "Cross-Industry Standard Process for Data Mining" beschreibt den Standard für die Strukturierung der Maschinellen Lernalgorithmen. Das Ziel ist es, einen einheitlichen Prozess für Data Mining Projekte zu schaffen und ist eine Anleitung für Data Mining, die in 6 Schritten aufgeteilt ist [Wut19]. Der Ablauf des Modells ist in Abbildung 3.1 dargestellt.

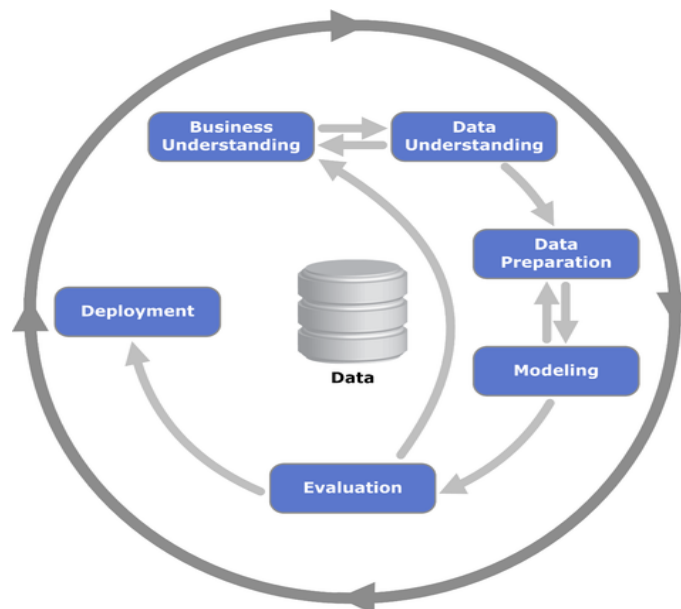


Abbildung 3.1: Ablauf CRIPS-DM Modell [Abbb]

Business Understanding

Der erste Schritt der 6 Schritte des CRISP-DM Modells, ist das Business Understanding, welches sich mit den betriebswirtschaftlichen Problemen befasst. Zunächst werden betriebswirtschaftliche Anforderungen extrahiert und analysiert, wodurch die Grundlagen für weitere Schritte und Entscheidungen gebildet werden. Damit ein Verständnis über die betriebswirtschaftlichen Fragestellungen übermittelt wird, ist es wichtig die Anwender in die Data Mining Prozesse mit einzubeziehen.

Das Business Understanding beinhaltet 4 Aspekte, die in Betracht gezogen werden sollen. Hierzu gehört wie zuvor auch erwähnt, die Bestimmung der betriebs-

wirtschaftlichen Problemstellungen und dessen Zielkriterien. Ein weiterer Aspekt, ist die Situationsbewertung, welches die zur Verfügung gestellten Ressourcen für das Data Mining Projekt beschreibt. Dies bedeutet zum Beispiel die Software- und Personalressourcen. Zusätzlich werden Risiken, die, während dem Data Mining Projekt entstehen können, identifiziert. Auch die Bestimmung analytischer Ziele ist ein Aspekt des Business Understanding, indem eine Analyse der zuvor bestimmten betriebswirtschaftlichen Probleme durchgeführt wird. Auch Erfolgskriterien für das Data Mining Projekt werden bestimmt, wie zum Beispiel eine Optimierung des Geschäftsprozesses um 5% mit weniger Bedarf an Ressourcen. Der letzte Aspekt beinhaltet die Erstellung des Projektplans, welches die beabsichtigten Ziele des Projektes beschreibt, indem die Auflistung der einzelnen Schritte mit jeweiliger Zeitspanne definiert werden. Hinzu kommen die möglichen Risiken oder Ursachen für das Scheitern des Projektes, die aufgefasst werden.

Data Understanding

In diesem Schritt werden relevante Datenbestände, die zu einer Hypothese führen können, aufgestellt. Zum Beispiel werden bei einer Wettervorhersage die aktuellen Zustände der Atmosphäre betrachtet. In diesem Fall ist der Datenbestand die Zustände der Atmosphäre. Zunächst werden Daten gesammelt, die benötigt werden und zu einer bereits bestehenden Datenmenge integriert. Als nächstes werden diese Datenbestände beschrieben. Hierzu werden Eigenschaften wie Quantität, Formateigenschaften, Anzahl der Einträge und Felder der Daten erfasst. Der letzte Schritt des Data Understanding ist die Untersuchung und die Bewertung der Daten, in dem eine Analyse mit den Daten betrieben wird, um erste Hypothesen und Erkenntnisse zu treffen und diese dann zu dokumentieren. Bei der Bewertung der Daten wird sichergestellt, dass die Daten verwendbar sind und sie keine fehlende Attributwerte enthalten [Wut19].

Data Preparation

Um die finale Menge der Datenauswahl zu erstellen, wird der Schritt Data Preparation durchgeführt. An erster Stelle werden Daten ausgewählt, die für das Ziel des Projektes relevante Informationen hergeben. Wichtig hierbei ist zu betrachten, dass die Daten in dem richtigen Datenformat gespeichert werden [Wut19].

Modeling

Um eine vorliegende Problemstellung aus dem Bereich Maschinelles Lernen lösen zu können, stehen verschiedene Modelle beziehungsweise Algorithmen zur Verfügung.

Aus diesen Modellen wird dann das Model ausgesucht, welches im Bezug zum definierten Ziel das optimale Ergebnis liefert. Manche Modelle benötigen hierzu vorbestimmte Datenstrukturen, die dazu führen können, einen Schritt zurück zu gehen und sich noch einmal mit der Datenvorbereitung zu befassen. Anschließend wird das Modell trainiert und getestet um die Qualität und Genauigkeit des Modells zu überprüfen [Wut19].

Evaluation

Vor der Finalisierung des Data Mining Projektes, ist es noch notwendig zu prüfen, ob die Qualität des Modelles für das Ziel des Projektes ausreicht. Ist das Ergebnis noch nicht gut genug oder werden vorher definierte Ziele nicht erreicht, müssen die Gründe für das Scheitern analysiert und entsprechende Änderungen im Modell vorgenommen werden. Dies hat zur Folge, dass die vorherigen Schritte nochmals durchlaufen werden müssen [Wut19].

Deployment

Als letztes werden die aus dem erstellten Modell gewonnen Erkenntnisse für die Nutzer vorbereitet und ihm als Anwendung zur Verfügung gestellt. Damit kann ein Anwender zum Beispiel eigene Daten in das Modell hochladen und erhält, je nach der vorliegenden Problematik, ein Ergebnis für seine verwendeten Daten [Wut19].

4 Experimente

4.1 Business Understanding

Online-Versandhändler haben oft eine hohe Rückläuferquote. Diese Rückläuferquote kommt zum Beispiel durch den Kauf von Artikeln in verschiedenen Größen zustande. Die nicht passenden Größen werden dann vom Käufer wieder zurückgesendet. Unter diese Versandhändler fällt, mit einer Rückläuferquote von bis zu 50%, auch das Unternehmen Zalando. Dabei müssen 97% der zurückgesendeten Produkte wieder eingelagert und verkauft werden [C.18]. Für den erneuten Verkauf werden die Produkte, in diesem Betrachtungsfall Klamotten, identifiziert, gelabelt und bei Bedarf wieder eingelagert. Das Unternehmen Zalando verschickte im Jahr 2019 rund 186 Millionen Bestellungen [Raber]. Wenn man davon ausgeht, dass eine Bestellung 4 Artikel beinhaltet und 50% der Bestellungen wieder zurückgesendet werden, ergibt das 372 Millionen Artikel, die wieder neu gelabelt und eingelagert werden müssen.

Die Unterstützung und Optimierung der Rückläuferbearbeitung, kann zum Beispiel durch eine AI gestützte Klassifikation gewährleistet werden. Mit diesem Ansatz kann dann die zu einem Artikel zugehörige Kategorie, auf Basis von Bildern, bestimmt werden. Damit ergibt sich die Frage- bzw. Problemstellung, inwiefern können Machine Learning Algorithmen bei der Klassifizierung von Bildern einzelner Kleidungsstücke unterstützen?

4.2 Data Understanding

Der für diese Arbeit gewählte Datensatz heißt Fashion MNIST und wurde im Jahr 2017 von dem Unternehmen Zalando veröffentlicht. Der Datensatz steht in einer GitHub Repository zur Verfügung und ist frei zugänglich [SE17]. Er beinhaltet Zalando-Artikelbilder, die in zwei Datensätze aufgeteilt sind. Der erste Satz enthält 60.000 Beispiele, die für das Training verwendet werden und der zweite Satz 10.000 Beispiele, die für das Testen verwendet werden. Jedes Beispiel, welches jeweils ein Bild darstellt, ist ein 28x28 Graustufenbild. Jedem Bild wird dabei noch ein Label zugeordnet. Ein Label repräsentiert eins der 10 Klassen, wobei jede Klasse eine

Artikelkategorie darstellt. In der folgenden Abbildung 4.1 ist eine Übersicht des Datensatzes bzw. der Bilder zu sehen. Dabei werden pro Klasse drei Zeilen beansprucht.

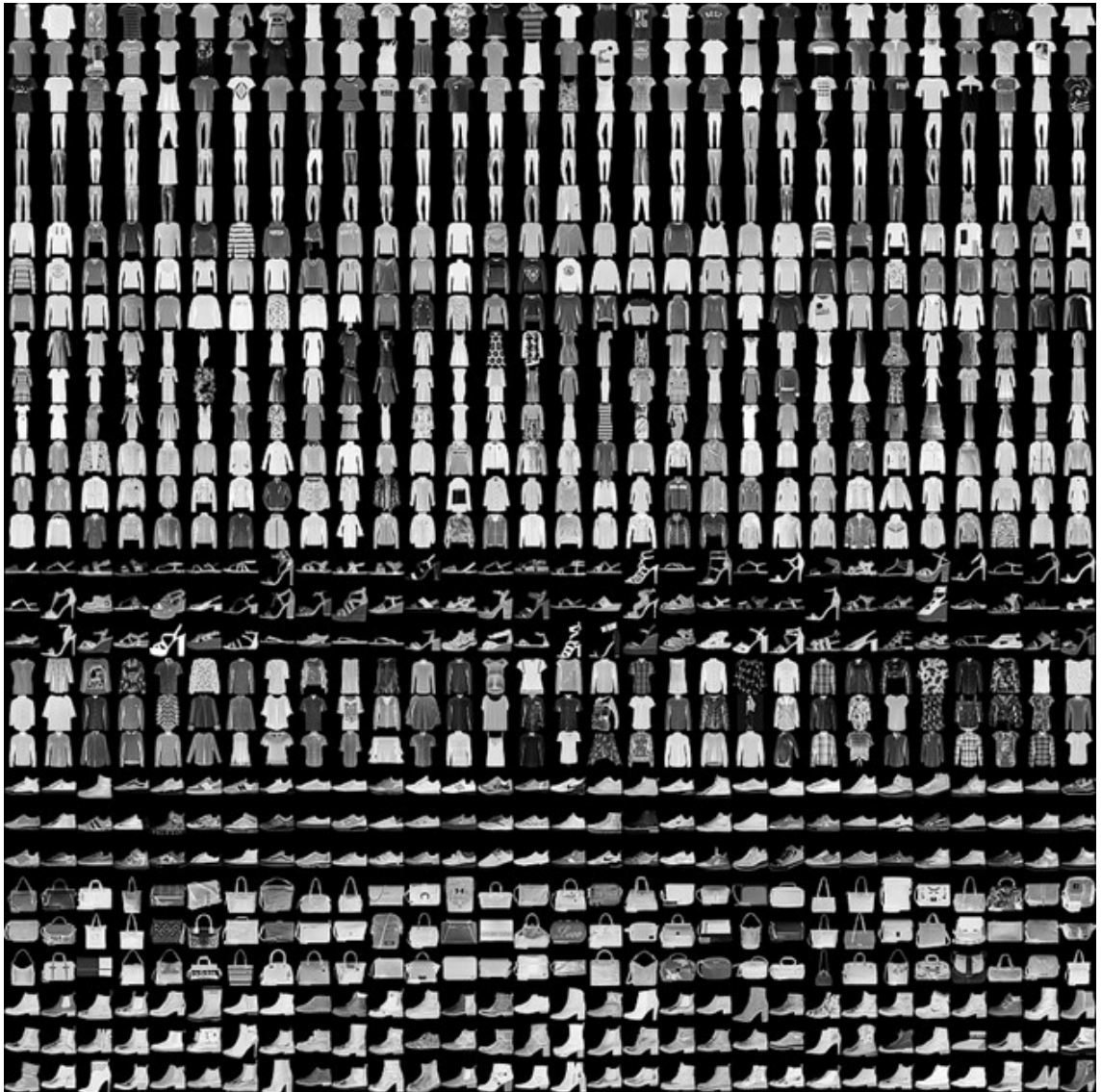


Abbildung 4.1: Der gesamte Datensatz: FASHION-MNIST [Abbc]

Der ursprüngliche MNIST Datensatz, der durch Fashion-MNIST ersetzt wurde, ist sehr beliebt und wird oft verwendet, um Algorithmen zu validieren. Es heißt auch „If it doesn’t work on MNIST, it won’t work at all”. Jedoch wurde der MNIST Datensatz aufgrund seiner Einfachheit ausgetauscht [SE17].

Auch die Labels bestehen aus 28x28 Pixeln, wobei jeder einzelne Pixel einen Wert zwischen 0 und 255 annehmen kann. Der aktuelle Datensatz Fashion-MNIST enthält folgende 10 Klassen bzw. Labels:

Label	Artikel
0	T-shirt/top
1	Trouser
2	Pullover
3	Dress
4	Coat
5	Sandal
6	Shirt
7	Sneaker
8	Bag
9	Ankle boot

Tabelle 4.1: Labels des Datensatzes

Der Trainings- als auch der Testdatensatz ist ziemlich ausgeglichen. In jedem Satz ist jede Kategorie mit ca. 10% vertreten, was in den nächsten Abbildungen zu sehen ist.

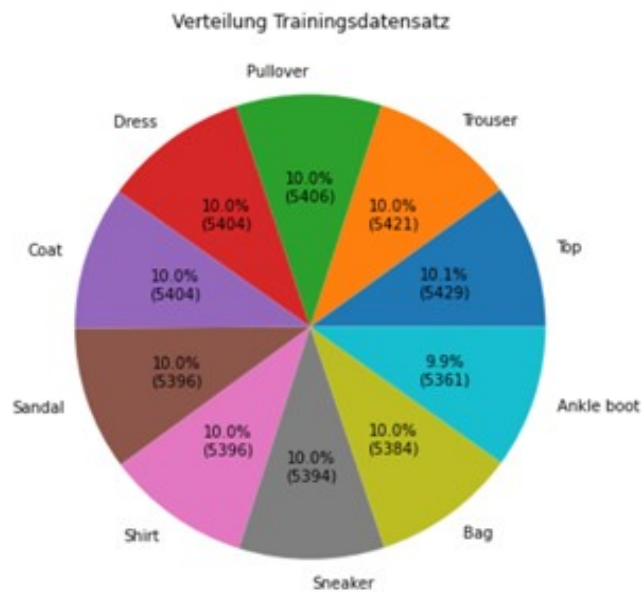


Abbildung 4.2: Verteilung Trainingsdatensatz [Abbh]



Abbildung 4.3: Verteilung Testdatensatz [Abbg]

Aufgrund der Ausgeglichenheit der Daten müssen im Rahmen der Data Preparation keine großen Änderungen oder Vorbereitungen vorgenommen werden.

4.3 Data Preparation

Um die vorliegenden Daten für die modellierten ML-Verfahren verwenden zu können, müssen diese entsprechend vorbereitet werden.

4.3.1 Test- und Trainingsdaten

Für die Vorbereitung werden die Trainingsdaten nochmal aufgeteilt. Dabei werden aus den vorhandenen 60000 Trainingssätzen, 6000 für die Validierung extrahiert. Dadurch bleiben für die Trainingsphase 54000 Datensätze übrig. Die Testdatensätze bleiben mit der ursprünglichen Anzahl von 10000 gleich.

4.3.2 Skalierung der Merkmale

Die Skalierung von Merkmalen dient zur Normalisierung des Bereichs der erklärenden bzw. unabhängigen Variablen. Das Verfahren ist auch als Datennormalisierung bekannt und wurde im Rahmen der Datenvorbereitung auf den Daten angewendet. Die Eingabedaten liegen als Graustufenbilder vor und besitzen Werte zwischen 0 und 255. Um die Normalisierung dieser Daten vorzunehmen, werden die Werte durch 255 geteilt um somit Ausprägungen zwischen 0 und 1 zu erreichen.

4.3.3 Datentransformation

In dem Ausgangsdatensatz stellt jede Zeile mit seinen 784 Spalten, die wiederum als 28x28 Pixel zu verstehen sind, jeweils ein Bild dar. Für die spätere Verwendung der Daten, müssen diese aber umgeformt werden. Die Umformung der Graustufenbilder muss so erfolgen, dass ein Bild durch 28 Zeilen und 28 Spalten dargestellt wird und so die Pixel wiedergibt. Dadurch kann das Neuronale Netz eine genaue Position der Pixel identifizieren. Die dafür verwendete Methode kann dem beigefügten Notebook entnommen werden. Im Rahmen der Datentransformation werden außerdem die Zielvariablen in die vorhandenen 10 Kategorien unterteilt.

4.4 Modellierung

Insgesamt werden drei verschiedene implementierte Modelle vorgestellt und im nächsten Abschnitt evaluiert. Jedes Modell bzw. Experiment beinhaltet dabei ein Convolutional Neuronal Network. Die Unterschiede der drei Modelle liegen dabei in den verwendeten Schichten und deren Zusammenspiel.

4.4.1 Experiment 1

Das erste implementierte CNN-Modell beinhaltet folgende drei Layer, die in vorherigen Abschnitten bereits erläutert wurden:

- Convolutional Layer (Conv2D)
- Pooling Layer (MaxPooling2D)
- Flatten Layer (Flatten)
- Fully-Connected Layer (Dense)

Der Conv2D und MaxPooling2D Layer gehören, wie in Kapitel 2 beschrieben, zu dem Feature-Extraktor. Die beiden Schichten werden zwei mal nacheinander in das Modell integriert. Im Anschluss werden dem Modell zwei Dense Layer hinzugefügt. Der erste Conv2D Layer bekommt die Eingabedaten im 28x28 Pixel Format, daher erhält diese Schicht einen Input-Parameter, der die Dimension der Eingabedaten beschreibt. Im MaxPooling2D werden die aus den Conv2D extrahierten Merkmale in eine 14x14 dimensionale Matrix zusammengefasst. Der zweite Conv2D Layer erhält nach dem MaxPooling2D Layer Daten in der Form 14x14 Pixeln, aus dem der Layer dann wiederum Merkmale extrahiert. Der anschließende MaxPooling2D Schicht fasst diese Merkmale dann nochmals zusammen. Als nächstes formt der Flatten Layer die Merkmale in eine eindimensionale Matrix um. Zuletzt folgen dann die beiden Dense

Layer. Die erste Dense Schicht ist der Fully-Connected Layer und die zweite eingesetzte Dense Schicht der Output Layer (Ausgabeschicht). Der ersten Dense Schicht wird dabei die Aktivierungsfunktion Rectifier übergeben. Die Rectifier (relu) wird wie folgt definiert.

$$f(x) = \max(0, x), \text{ wobei } x \text{ der Eingangswert ist.}$$

Sie beschreibt seinen Argument als Positivteil. Für die Ausgabeschicht wird die Aktivierungsfunktion softmax eingestellt. Damit wird das erste Modell finalisiert und kann trainiert werden.

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 32)	320
max_pooling2d (MaxPooling2D)	(None, 14, 14, 32)	0
conv2d_1 (Conv2D)	(None, 14, 14, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 7, 7, 32)	0
flatten (Flatten)	(None, 1568)	0
dense (Dense)	(None, 28)	43932
dense_1 (Dense)	(None, 10)	290

```

=====
Total params: 53,790
Trainable params: 53,790
Non-trainable params: 0
=====

```

Abbildung 4.4: Experiment 1: Modellaufbau [HK221] (Experiment 1)

Im gesamten kann das Model 53,790 Parameter trainieren. Bevor das Training gestartet werden kann, müssen noch einige Variablen definiert werden. Zunächst braucht das Model ein optimizer der ein tensorflow Objekt darstellt. Dieses Objekt beschreibt den Optimierung Vorgang für das Modell. In diesem Fall wird der optimizer als "tf.keras.optimizers.Adam()" definiert, der stochastisches Gradientenabstiegsverfahren beschreibt. Der nächste Parameter definiert die Verlustfunktion des Modells, welches wir als Cross-Entropy-Loss Funktion angeben. Als Bewertungsmetrik wird dem Model die Genauigkeit (engl. accuracy) ausgewählt, die Prozentual angibt wie gut ein Model prognostiziert. Zu letzte werden Epochen anzahl und Batch größen angegeben.

Trainingsphase

Mit dem Fitten der Trainings- und Validierungsdaten wird die Trainingsphase gestartet. Nachdem die Trainingsphase beendet ist werden Bewertungsmetriken genutzt um eine Darstellung der Leistung des Modells zu bekommen.

```
Epoch 1/10
844/844 [=====] - 17s 20ms/step - loss: 0.5194 - accuracy: 0.8144 - val_loss: 0.3874 - val_accuracy: 0.8656
Epoch 2/10
844/844 [=====] - 17s 20ms/step - loss: 0.3344 - accuracy: 0.8802 - val_loss: 0.3154 - val_accuracy: 0.8813
Epoch 3/10
844/844 [=====] - 17s 20ms/step - loss: 0.2968 - accuracy: 0.8934 - val_loss: 0.3070 - val_accuracy: 0.8797
Epoch 4/10
844/844 [=====] - 17s 20ms/step - loss: 0.2709 - accuracy: 0.9025 - val_loss: 0.3085 - val_accuracy: 0.8906
Epoch 5/10
844/844 [=====] - 17s 20ms/step - loss: 0.2520 - accuracy: 0.9087 - val_loss: 0.2713 - val_accuracy: 0.9031
Epoch 6/10
844/844 [=====] - 17s 21ms/step - loss: 0.2310 - accuracy: 0.9156 - val_loss: 0.2803 - val_accuracy: 0.9031
Epoch 7/10
844/844 [=====] - 17s 20ms/step - loss: 0.2159 - accuracy: 0.9211 - val_loss: 0.3079 - val_accuracy: 0.8828
Epoch 8/10
844/844 [=====] - 17s 20ms/step - loss: 0.2009 - accuracy: 0.9269 - val_loss: 0.2420 - val_accuracy: 0.9141
Epoch 9/10
844/844 [=====] - 17s 20ms/step - loss: 0.1882 - accuracy: 0.9310 - val_loss: 0.2347 - val_accuracy: 0.9187
Epoch 10/10
844/844 [=====] - 17s 20ms/step - loss: 0.1766 - accuracy: 0.9343 - val_loss: 0.2497 - val_accuracy: 0.9141
```

Abbildung 4.5: Experiment 1: Bewertung Trainingsphase [HK221](Experiment 1)

Die Ergebnisse zeigen, dass die erste Epoche eine Präzision von 86,56% hat und die sich bei den nächste Epoche erhöhen. Damit hat das Modell am Ende eine Präzision während dem Training von 91%. Sichtbar ist die Differenz zwischen *accuracy* (93,34%) und *val_accuracy* (91,14%). Die *accuracy* (auch Genauigkeit genannt) beschreibt in diesem Fall, die Genauigkeit der Trainingsdaten, diese wird jedoch nicht für die Bewertung genutzt. Die *val_accuracy*, welches die Genauigkeit der Validierungsdaten anzeigt wird hier in Betracht gezogen, da diese zur Validierung des Modells erstellt worden ist und komplett abgekapselt von den Trainingsdaten übergeben wurde. Der nächste Schritt zur Optimierung des Modells sollte die Anpassung sein, welches den Abstand der Genauigkeiten verringert. Es ist auch ersichtlich, dass die *loss*-Werte gegen null näher. Dies zeigt, dass das Modell sich verbessert bzw. optimiert.

4.4.2 Experiment 2

In diesem Experiment wird das Modell aus dem Experiment 1 übernommen und weiter verbessert. Das Ziel ist zunächst die Reduzierung der Differenz von *val_accuracy* und *accuracy*, damit das Modell genauere Bewertungen liefern kann.

Literaturverzeichnis

- [Abba] *Aufbau eines CNN.* https://www.researchgate.net/figure/Allgemeiner-Aufbau-eines-CNN-In-Anlehnung-an-MathWorks-2017_fig4_342750457, Zuletzt aufgerufen am 12.11.2021.
- [Abbb] *CRISP-DM Modell.* https://statistik-dresden.de/wp-content/uploads/2012/04/CRISP-DM_Process_Diagram1.png, Zuletzt aufgerufen am 12.11.2021.
- [Abbc] *FASHION-MNIST.* <https://github.com/zalandoresearch/fashion-mnist/blob/master/doc/img/fashion-mnist-sprite.png>, Zuletzt aufgerufen am 12.11.2021.
- [Abbd] *KNN vs. DL.* <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0192-5>, Zuletzt aufgerufen am 12.11.2021.
- [Abbe] *Modell eines Neurons.* https://datasolut.com/wp-content/uploads/2019/11/neuronale_netze_aktivierung-1024x485.png, Zuletzt aufgerufen am 12.11.2021.
- [Abbf] *Schichten der KNN.* <https://datasolut.com/wp-content/uploads/2019/10/k%C3%9Cnstliche-neuronale-Netze.jpg>, Zuletzt aufgerufen am 12.11.2021.
- [Abbg] *Verteilung Testdatensatz.* <https://github.com/zalandoresearch/fashion-mnist>, Zuletzt aufgerufen am 12.11.2021.
- [Abbh] *Verteilung Trainingsdatensatz.* <https://github.com/zalandoresearch/fashion-mnist>, Zuletzt aufgerufen am 12.11.2021.
- [C.18] Hoefer C. Wie amazon, zalando und otto mit retouren umgehen. *manager magazin*, 2018. <https://www.manager-magazin.de/unternehmen/handel/retouren-wie-amazon-zalando-und-otto-mit-ruecksendungen-umgehen-a-1213.html>, Zuletzt aufgerufen am 13.11.2021.

- [Cho18] François Chollet. *Deep Learning mit Python und Keras : Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. [Frechen] : mitp Verlags, 2018.
- [FG18] Fraunhofer-Gesellschaft. Maschinelles lernen eine analyse zu kompetenzen,forschung und anwendung. *Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V*, 2018. https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publicationen/Fraunhofer_Studie_ML_201809.pdf, Zuletzt aufgerufen am 22.02.2022.
- [Fro21] Jörg Frochte. *Maschinelles Lernen : Grundlagen und Algorithmen in Python*. Hanser eBook, 2021.
- [Hal22] Halil. *Feed Forward Netz*, 2022. https://github.com/halil-kocak/Abschlussarbeit-ErkennungVonKleiderstuecken/blob/main/Dokumentation/FF_Netz.PNG?raw=true, Zuletzt aufgerufen am 21.02.2022.
- [HK221] *Jupyter Notebook: Erkennung von Kleiderstücken*, Halil Kocak 2021. https://github.com/halil-kocak/Abschlussarbeit-ErkennungVonKleiderstuecken/blob/main/Code/defaultCNN/default_cnn.ipynb, Zuletzt aufgerufen am 25.02.2021.
- [Lub19] Dipl.-Ing. Stefan Luber. Was ist ein convolutional neural network? *BIGDATA INSIDER*, 2019. <https://www.bigdata-insider.de/was-ist-ein-convolutional-neural-network-a-801246/>, Zuletzt aufgerufen am 12.11.2021.
- [Mah18] Hamza Mahmood. The softmax function, simplified. *Towards Data Science*, 2018. <https://towardsdatascience.com/softmax-function-simplified-714068bf8156>, Zuletzt aufgerufen am 22.02.2022.
- [Mat21] Stephan Matzka. *Künstliche Intelligenz in den Ingenieurwissenschaften*. Springer Fachmedien Wiesbaden, 2021.
- [Pla21] Matthias Plaue. *Data Science*. SpringerSpektrum, 2021.
- [Raber] L. Rabe. Jährliche anzahl der bestellungen bei zalando bis 2020. *Statista*, 2021, 10. November. <https://de.statista.com/statistik/daten/studie/325862/umfrage/>

`anzahl-aller-bestellungen-bei-zalando-pro-jahr/`, Zuletzt aufgerufen am 12.11.2021.

- [SE17] Zalando SE. Fashion-mnist [datensatz]. 2017. <https://github.com/zalando-research/fashion-mnist>, Zuletzt aufgerufen am 10.11.2021.
- [uSV21] Sandro Scheid und Stefanie Vogl. *Data Science*. Carl Hanser Verlag GmbH und Co. KG, 2021.
- [Wut19] Laurenz Wuttke. Crisp-dm: Grundlagen, ziele und die 6 phasen des data mining prozess. *datasolut*, 2019. <https://datasolut.com/crisp-dm-standard/>, Zuletzt aufgerufen am 10.11.2021.