

## Klassifikation von Modebildern mit Fashion-MNIST

Mia Horvat

WWI2021F, DHBW Stuttgart

### ABSTRACT:

In diesem Projekt wurde ein Convolutional Neural Network (CNN) zur Klassifikation von Modebildern aus dem Fashion-MNIST-Datensatz entwickelt. Der Datensatz umfasst 70.000 Bilder in verschiedenen Graustufen von Kleidungsstücken, die in zehn Kategorien unterteilt sind. Ziel war es, ein Modell zu entwickeln, das nicht nur präzise klassifiziert, sondern auch zur Outfiterstellung genutzt werden kann. Die Daten wurden normalisiert und durch Techniken wie Rotationen und Spiegelungen erweitert, um Overfitting zu reduzieren. Ein Grid Search wurde für das Hyperparameter-Tuning eingesetzt, wobei der Adam-Optimierungsalgorithmus zur Anwendung kam. Das Modell erreichte eine Genauigkeit von 93% auf dem Testdatensatz und zeigte eine gute Balance zwischen Präzision und Wiedererkennung. Die erfolgreiche Outfiterstellung demonstrierte die praktische Anwendbarkeit des Modells. Der vollständige Quellcode und weitere Informationen sind im GitHub-Repository verfügbar, dessen Link am Ende dieses Dokuments zu finden ist.

### 1. INTRODUCTION

In diesem Projekt geht es um die Klassifikation von Modebildern mit dem Fashion-MNIST-Datensatz (siehe Abbildung 1). Fashion-MNIST ist eine bekannte Datenbank, die oft für das Training von Bildverarbeitungssystemen genutzt wird. Sie enthält 70.000 Graustufenbilder von Kleidungsstücken, die jeweils einer von zehn Kategorien zugeordnet sind, darunter T-Shirts, Hosen, Pullover, Kleider, Mäntel, Sandalen, Hemden, Sneaker, Taschen und Stiefeletten.

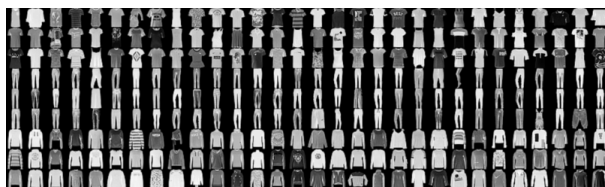


Abbildung 1: Überblick der Vielfältigkeit des Datensatzes

Der Datensatz ist in zwei Teile gegliedert:

- **Trainingsset:** 60.000 Bilder zur Modellbildung.
- **Testset:** 10.000 Bilder zur Modellvalidierung.

Jedes Bild hat eine Auflösung von 28x28 Pixeln, was insgesamt 784 Pixelwerte ergibt. Diese Merkmale repräsentieren die Pixelintensitäten des Bildes.

Die Klassifikation von Modebildern kann besonders herausfordernd sein, da viele Kleidungsstücke visuell sehr ähnlich sind und es innerhalb einer Kategorie große Unterschiede geben kann.

Der Datensatz war bereits gut vorbereitet und erforderte keine umfangreiche Datenbereinigung. Eine Analyse des Datensatzes zeigte keine fehlenden Werte, und die Daten waren über alle Klassen hinweg ausgewogen. Daher waren keine zusätzlichen Bereinigungsschritte notwendig.

**Ziel des Projekts:** Das Hauptziel dieses Projekts ist es, ein präzises und robustes Modell zur Klassifikation von Modebildern zu entwickeln. Dabei sollen nicht nur die Kleidungsstücke korrekt erkannt werden, sondern das Modell soll auch in der Lage sein, basierend auf den Klassifikationsergebnissen Outfits zusammenzustellen. Dies bietet praktische Anwendungen, beispielsweise bei der Unterstützung von Online-Shopping-Erlebnissen.

**Datenaufbereitung:** Zur Datenaufbereitung wurden die Pixelwerte normalisiert, um die Trainingszeit zu verkürzen und die Stabilität des Modells zu verbessern. Zudem wurden Techniken wie Dataaugmentation angewendet, um die Robustheit des Modells zu erhöhen und Overfitting zu reduzieren. Die Trainings- und Testdaten wurden aus CSV-Dateien geladen und entsprechend vorbereitet.

## 2. FEATURE ENGINEERING

Feature Engineering ist entscheidend für die Verbesserung der Modellleistung. Im Rahmen dieses Projekts wurden folgende Schritte durchgeführt:

Die Normalisierung der Pixelwerte hilft dabei, die Stabilität des Modells während des Trainings zu verbessern und die Trainingszeit zu verkürzen. Die Pixelwerte werden von  $[0, 255]$  auf  $[0, 1]$  skaliert, indem jeder Pixelwert durch 255 dividiert wird:

```
train_images = train_images / 255.0
```

```
test_images = test_images / 255.0
```

Die Bilder im Fashion-MNIST-Datensatz haben ursprünglich eine flache Struktur mit 784 Pixelwerten (28x28 Pixel). Um sie als Eingabe für das CNN vorzubereiten, mussten die Bilder in ein 4D-Format

umgewandelt werden. Dieser Schritt ist notwendig, da CNNs in der Lage sind, räumliche Hierarchien in den Bilddaten zu lernen und effizient zu verarbeiten. Das Umformen der Daten in ein 4D-Format (28x28x1) ermöglicht es dem Netzwerk, die Bilder als zweidimensionale Gitter mit einer Kanal-Tiefe von 1 zu betrachten, was die Merkmalsextraktion durch die Convolutional Layers erleichtert:

```
train_images =  
train_images.reshape((train_images.shape  
[0], 28, 28, 1))
```

```
test_images =  
test_images.reshape((test_images.shape[0]  
, 28, 28, 1))
```

Dataaugmentation erhöht die Vielfalt der Trainingsdaten, was dazu beiträgt, Overfitting zu reduzieren und die Robustheit des Modells zu erhöhen. Techniken wie zufällige Rotation, Zoom und horizontale Spiegelung wurden angewendet:

```
datagen = ImageDataGenerator(  
    rotation_range=10,  
    zoom_range=0.1,  
    horizontal_flip=True)  
datagen.fit(train_images)
```

Durch diese Schritte wurde die Qualität und Vielfalt der Eingabedaten verbessert, was die Grundlage für die erfolgreiche Klassifikation durch das neuronale Netzwerk bildet.

## 3. FORMULIERUNG DER ZIELMETRIK

Für die Klassifikation von Modebildern im Fashion-MNIST-Datensatz wurde die "Genauigkeit" (Accuracy) als Zielmetrik verwendet. Diese Metrik misst den Anteil

der korrekt klassifizierten Bilder an der Gesamtzahl der Bilder und ist eine weit verbreitete Methode zur Bewertung von Klassifikationsproblemen.

Um ein umfassenderes Bild der Modellleistung zu erhalten, wurden neben der Genauigkeit auch weitere Metriken herangezogen. Die Präzision (Precision) gibt an, wie viele der als positiv vorhergesagten Fälle tatsächlich positiv sind. Eine hohe Präzision bedeutet, dass das Modell wenige falsch-positive Vorhersagen macht. Der Recall, auch als Wiedererkennungsrates bekannt, misst den Anteil der korrekt identifizierten positiven Fälle unter allen tatsächlichen positiven Fällen. Diese Metrik zeigt, wie gut das Modell darin ist, tatsächliche positive Fälle zu erkennen.

Um eine Balance zwischen Präzision und Recall zu finden, wurde der F1-Score berechnet. Dieser stellt das harmonische Mittel von Präzision und Recall dar und bietet ein ausgewogenes Maß für die Modellleistung.

Durch die Verwendung dieser verschiedenen Metriken konnte die Leistung des Modells umfassend bewertet werden. Dies stellte sicher, dass das Modell nicht nur eine hohe Genauigkeit aufweist, sondern auch zuverlässig und präzise bei der Klassifikation der Modebilder ist.

#### 4. FORMULIERUNG DES NEURONALEN NETZES

Für die Klassifikation von Modebildern im Fashion-MNIST-Datensatz wurde ein Convolutional Neural Network (CNN) entworfen. Das Netzwerk beginnt mit einer Eingabeschicht, die die normalisierten und umgeformten Bilder (28x28x1) aufnimmt. Im Anschluss daran folgen mehrere

Convolutional-Schichten, die mithilfe von Aktivierungsfunktionen relevante Merkmale aus den Bildern extrahieren. Diese Schichten verwenden Filter, um verschiedene räumliche Strukturen und Muster in den Bildern zu erkennen.

Um die Datenmenge zu reduzieren und die wichtigsten Merkmale hervorzuheben, werden Pooling-Schichten eingesetzt. Max-Pooling, mit einer Poolgröße von 2x2, hilft dabei, die Anzahl der Parameter zu verringern und eine Überanpassung zu vermeiden, indem es die räumlichen Dimensionen der Merkmalskarten reduziert.

Zwischen den Convolutional-Schichten befinden sich Dropout-Schichten, die während des Trainings zufällig einige Neuronen deaktivieren. Diese Maßnahme trägt dazu bei, die Robustheit des Modells zu erhöhen und eine Überanpassung zu verhindern. Die Dropout-Rate beträgt 0.25 nach jeder Convolutional-Schicht und 0.5 nach der vollständig verbundenen Schicht.

Nach den Convolutional- und Pooling-Schichten folgt eine Flatten-Schicht, die die zweidimensionalen Merkmalskarten in einen eindimensionalen Vektor umwandelt. Dieser Vektor dient als Eingabe für die vollständig verbundenen Schichten des Netzwerks. Die vollständig verbundenen Schichten bestehen aus 128 Neuronen mit ReLU-Aktivierungsfunktionen, welche es dem Modell ermöglichen, komplexe Beziehungen in den extrahierten Merkmalen zu lernen.

Abschließend liefert die Ausgangsschicht die endgültigen Klassifikationswerte für jede Kategorie. Diese Schicht verwendet die Softmax-Aktivierungsfunktion, um Wahrscheinlichkeiten für jede der zehn

Klassen im Datensatz zu berechnen. Zur Optimierung des Modells wurde der Adam-Optimierungsalgorithmus verwendet, da sie sich besonders gut für Mehrklassen-Klassifikationsprobleme eignet.

Durch diese Architektur wurde ein robustes und genaues Modell zur Klassifikation der Modebilder im Fashion-MNIST-Datensatz entwickelt.

## 5. HYPERPARAMETER-TUNING

Das Hyperparameter-Tuning wurde durchgeführt, um die Leistung des CNN zu maximieren. Dabei wurden die wichtigsten Hyperparameter angepasst, darunter die Lernrate, die Batchgröße und die Anzahl der Epochen. Die optimalen Werte wurden mit einer Lernrate von 0.0004156711774145427, einer Batchgröße von 32 und 30 Epochen ermittelt.

Zur Optimierung dieser Hyperparameter wurde der Adam-Optimierungsalgorithmus eingesetzt, da er für seine effiziente Anpassung der Lernrate bekannt ist. Um die besten Kombinationen dieser Hyperparameter zu finden, wurde die Methode der Grid Search verwendet. Diese systematische Suchmethode ermöglicht es, alle möglichen Kombinationen innerhalb eines festgelegten Hyperparameter-Raums zu prüfen.

Um Overfitting zu reduzieren und die Robustheit des Modells zu erhöhen, wurden verschiedene Techniken der Dataaugmentation angewendet. Dazu gehörten zufällige Rotationen, Zooms und horizontale Spiegelungen, die den Datensatz künstlich erweiterten und so die Generalisierungsfähigkeit des Modells verbesserten.

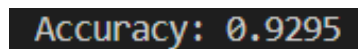
Die Validierung des Modells erfolgte durch die Überwachung eines Validierungssatzes. Frühzeitiges Stoppen (Early Stopping) wurde eingesetzt, um das Training zu beenden, sobald die Validierungsgenauigkeit über mehrere Epochen hinweg keine Verbesserung mehr zeigte. Diese Maßnahme verhinderte, dass das Modell überangepasst wurde.

Das Ergebnis des Hyperparameter-Tunings war eine verbesserte Modellgenauigkeit und eine effizientere Trainingszeit.

## 6. ERGEBNISSE UND FAZIT

Die Ergebnisse des Projekts belegen die Effektivität des entwickelten CNN bei der Klassifikation von Modebildern aus dem Fashion-MNIST-Datensatz.

Das Modell erzielte eine Genauigkeit von 93% auf dem Testdatensatz (siehe Abbildung 1).



Accuracy: 0.9295

Abbildung 2: Modellgenauigkeit

Diese hohe Genauigkeit zeigt, dass das CNN in der Lage ist, die meisten Kleidungsstücke korrekt zu klassifizieren.

Zusätzlich zur Genauigkeit wurden weitere Leistungsmetriken berechnet, um die Modellleistung umfassender zu bewerten. Die Präzision des Modells beträgt 92%, der Recall liegt bei 93% und der F1-Score beträgt ebenfalls 92% (siehe Abbildung 3).

Classification Report:				
	precision	recall	f1-score	support
0	0.88	0.89	0.89	1000
1	0.99	0.99	0.99	1000
2	0.92	0.87	0.90	1000
3	0.92	0.93	0.93	1000
4	0.91	0.88	0.89	1000
5	0.99	0.98	0.99	1000
6	0.77	0.82	0.79	1000
7	0.97	0.96	0.97	1000
8	0.98	1.00	0.99	1000
9	0.96	0.98	0.97	1000
accuracy			0.93	10000
macro avg	0.93	0.93	0.93	10000
weighted avg	0.93	0.93	0.93	10000

Abbildung 3: Ergebnisse der Leistungsmetriken

Diese Metriken verdeutlichen, dass das Modell nicht nur eine hohe Genauigkeit aufweist, sondern auch eine ausgewogene Balance zwischen Präzision und Wiedererkennung erreicht.

Die Konfusionsmatrix bietet eine detaillierte Ansicht der Klassifikationsergebnisse, indem sie die Anzahl der korrekt und inkorrekt klassifizierten Bilder für jede Kategorie darstellt (siehe Abbildung 4).

Confusion Matrix:										
[	891	1	9	12	1	1	82	0	3	0]
[	0	988	1	9	0	0	2	0	0	0]
[	13	0	872	11	32	0	68	0	4	0]
[	14	7	9	932	16	0	21	0	1	0]
[	0	1	26	19	876	0	77	0	1	0]
[	0	0	0	0	0	981	0	10	2	7]
[	88	1	29	24	35	0	818	0	5	0]
[	0	0	0	0	0	6	0	962	0	32]
[	1	0	0	1	0	1	1	0	996	0]
[	0	0	0	0	0	1	0	19	1	979]

Abbildung 4: Konfusionsmatrix

Beispielsweise wurden von 1000 Stiefeletten 979 korrekt als Stiefeletten klassifiziert, während 21 falsch klassifiziert wurden. Diese detaillierte Analyse hilft dabei, die Stärken und Schwächen des Modells besser zu verstehen.

Ein weiteres bemerkenswertes Ergebnis des Projekts ist die Anwendung des Modells zur Outfiterstellung. Das Modell wurde erfolgreich genutzt, um Outfits basierend auf den

Klassifikationsergebnissen zusammenzustellen (siehe Abbildung 5).



Abbildung 5: Erstelltes Outfit

Dies demonstriert die praktische Anwendbarkeit des Modells, beispielsweise zur Unterstützung von Online-Shopping-Erlebnissen, indem passende Kleidungsstücke empfohlen werden.

Zusammenfassend bestätigen die Ergebnisse, dass das entwickelte CNN-Modell robust und präzise ist und sich gut für die Klassifikation von Modebildern eignet. Durch die erfolgreiche Anwendung zur Outfiterstellung wurde zudem die praktische Relevanz des Modells eindrucksvoll demonstriert.

Der vollständige Quellcode dieses Projekts ist auf GitHub verfügbar. Das Repository enthält wichtige Dateien wie das Jupyter-Notebook, das trainierte Modell, die Anforderungen und die Trainingshistorie. Besuchen Sie das Repository unter folgendem Link:

[https://github.com/MiaHorvat/Machine\\_Learning\\_Project.git](https://github.com/MiaHorvat/Machine_Learning_Project.git)