Hochschule Luzern HSLU

Artificial Intelligence & Machine Learning

Bachelorarbeit Simulation

Vergleich der Leistungsfähigkeit verschiedener Machine-Learning-Modelle zur Vorhersage von Zahlungsausfällen: Eine Analyse anhand des 'Default of Credit Card Clients'-Datensatzes

Verfasser:

Noel Gugler

Matrikel-Nummer: -

Dozent: -

Korreferent: -

Rotkreuz, 10. Dezember 2024

Inhaltsverzeichnis

[1 Einleitung 5](#_Toc186275387)

[1.1 Problemstellung 5](#_Toc186275388)

[1.2 Zielsetzung der Arbeit 5](#_Toc186275389)

[1.3 Aufbau der Arbeit 6](#_Toc186275390)

[2 Theoretische Grundlagen 7](#_Toc186275391)

[2.1 Kreditrisiko und Zahlungsausfälle 7](#_Toc186275392)

[2.2 Klassifikationsprobleme im maschinellen Lernen 8](#_Toc186275393)

[2.2.1 Definition und Grundlagen 8](#_Toc186275394)

[2.2.2 Algorithmische Ansätze 8](#_Toc186275395)

[2.2.3 Leistungsbewertung von Klassifikationsmodellen 9](#_Toc186275396)

[2.2.4 Herausforderungen und Limitationen 10](#_Toc186275397)

[2.3 Überblick über die untersuchten Machine-Learning-Modelle 10](#_Toc186275398)

[2.3.1 Logistische Regression 10](#_Toc186275399)

[2.3.2 Random Forest 11](#_Toc186275400)

[2.3.3 – XGBoost 12](#_Toc186275401)

[3 Datensatzbeschreibung und Explorative Datenanalyse 13](#_Toc186275402)

[3.1 Herkunft und Struktur des Datensatzes 13](#_Toc186275403)

[3.2 Beschreibung der Variablen 13](#_Toc186275404)

[3.3 Explorative Datenanalyse (EDA) 14](#_Toc186275405)

[3.3.1 Ziel der Analyse 14](#_Toc186275406)

[3.3.2 Verteilung der Zielvariablen 14](#_Toc186275407)

[3.3.3 Analyse der numerischen Variablen 14](#_Toc186275408)

[3.3.4 Analyse der kategorischen Variablen 14](#_Toc186275409)

[3.3.5 Korrelation zwischen Variablen 15](#_Toc186275410)

[4 Methodik 16](#_Toc186275411)

[4.1 Datenaufbereitung und Feature Engineering 16](#_Toc186275412)

[4.1.1 Ausreisseranalyse 16](#_Toc186275413)

[4.1.1.1 IQR-Methode 16](#_Toc186275414)

[4.1.1.2 Z-Score-Methode 17](#_Toc186275415)

[4.1.1.3 Visualisierung der Ergebnisse 17](#_Toc186275416)

[4.1.2 Entscheidung zur Ausreisserbehandlung 18](#_Toc186275417)

[4.1.3 Datenumwandlung 19](#_Toc186275418)

[4.1.3.1 Feature-Skalierung 19](#_Toc186275419)

[4.1.3.2 One Hot Encoding 20](#_Toc186275420)

[4.1.4 Feature Engineering 21](#_Toc186275421)

[4.1.4.1 Die Wahl der Verhältnisse: 21](#_Toc186275422)

[4.1.4.2 Nutzen der Verhältnisse 22](#_Toc186275423)

[4.2 Implementierung der Machine-Learning-Modelle 22](#_Toc186275424)

[4.2.1 Performancemetriken zur Bewertung der Modelle 22](#_Toc186275425)

[4.2.1.1 Accuracy, Precision, Recall, F1-Score 23](#_Toc186275426)

[4.2.1.2 ROC-Kurve und AUC-Wert 24](#_Toc186275427)

[4.2.1.3 Zusammenfassung 26](#_Toc186275428)

[4.2.2 XGBoost 26](#_Toc186275429)

[4.2.2.1 Cross-Validation Scores 26](#_Toc186275430)

[4.2.2.2 Test Accuracy 26](#_Toc186275431)

[4.2.2.3 Classification Report 27](#_Toc186275432)

[4.2.2.4 Anpassung der Klassengewichte 28](#_Toc186275433)

[4.2.2.5 Hyperparameter-Tuning 31](#_Toc186275434)

[4.2.2.6 Kreuzvalidierung 37](#_Toc186275435)

[4.2.2.7 ROC-Kurve für das XGBoost-Modell 39](#_Toc186275436)

[4.2.3 Random Forest 42](#_Toc186275437)

[4.2.4 Logistische Regression 42](#_Toc186275438)

[5 Ergebnisse und Diskussion 43](#_Toc186275439)

[5.1 Performancemetriken im Modellvergleich 43](#_Toc186275440)

[5.2 Stärken und Schwächen der untersuchten Modelle 43](#_Toc186275441)

[5.3 Praktische Relevanz der Ergebnisse 43](#_Toc186275442)

[6 Ausblick und Limitationen 44](#_Toc186275443)

[6.1 Herausforderungen bei der Implementierung in der Praxis 44](#_Toc186275444)

[6.2 Potenziale für zukünftige Arbeiten 44](#_Toc186275445)

[7 Fazit 45](#_Toc186275446)

[8 Literaturverzeichnis 46](#_Toc186275447)

[9 Anhang 47](#_Toc186275448)

[9.1 9.1 Python-Code für Datenverarbeitung und Modellimplementierung 47](#_Toc186275449)

[9.2 9.2 Zusätzliche Grafiken und Tabellen 47](#_Toc186275450)

[10 Mindmap Gruppenthema 49](#_Toc186275451)

[11 Dank 51](#_Toc186275452)

[12 Schlusserklärung 52](#_Toc186275453)

# Einleitung

## Problemstellung

In der modernen Finanzwelt spielen Kreditkarten eine zentrale Rolle im täglichen Leben von Millionen von Menschen. Gleichzeitig birgt die Vergabe von Krediten erhebliche Risiken für Kreditgeber, insbesondere durch Zahlungsausfälle, die zu finanziellen Verlusten führen können. Eine effektive Analyse und Vorhersage von Zahlungsausfällen ist daher entscheidend, um die Kreditvergabe zu optimieren und Risiken zu minimieren.

Im Zeitalter von Big Data und Künstlicher Intelligenz bieten Machine-Learning-Modelle eine vielversprechende Möglichkeit, dieses Problem anzugehen. Mit ihrer Fähigkeit, Muster in grossen Datenmengen zu erkennen, haben sie sich als wertvolle Werkzeuge für die Kreditrisikoanalyse erwiesen. Allerdings gibt es eine Vielzahl von Modellen, die in diesem Kontext eingesetzt werden können – von einfachen statistischen Methoden wie der Logistischen Regression bis hin zu komplexeren Algorithmen wie Random Forest und XGBoost. Jedes dieser Modelle bringt spezifische Stärken und Schwächen mit sich, und die Wahl des optimalen Modells hängt von zahlreichen Faktoren wie Datenstruktur, Modellinterpretierbarkeit und Leistung ab.

In dieser Arbeit wird der Fokus auf die Untersuchung der Leistungsfähigkeit verschiedener Machine-Learning-Modelle bei der Vorhersage von Zahlungsausfällen gelegt. Basierend auf dem Datensatz *"Default of Credit Card Clients"* aus Taiwan wird analysiert, wie gut die Modelle Logistische Regression, Random Forest und XGBoost Zahlungsausfälle vorhersagen können. Der Datensatz umfasst sowohl demografische Merkmale als auch Kredit- und Zahlungsinformationen von Kreditkartenkunden, was eine umfassende Analyse ermöglicht.

## Zielsetzung der Arbeit

Die Zielsetzung dieser Arbeit besteht darin, durch einen systematischen Vergleich die Stärken und Schwächen der Modelle zu bewerten und herauszufinden, welches Modell für diese Art von Klassifikationsproblem am besten geeignet ist. Dabei werden Performancemetriken wie Genauigkeit, Präzision, Recall, der F1-Score sowie die AUC der ROC-Kurve herangezogen, um die Vorhersageleistung der Modelle zu messen.

Die Ergebnisse dieser Arbeit sollen nicht nur einen Beitrag zur akademischen Forschung leisten, sondern auch praktische Implikationen für die Finanzindustrie aufzeigen, indem sie eine fundierte Grundlage für die Wahl geeigneter Machine-Learning-Modelle zur Kreditrisikoanalyse bieten.

## Aufbau der Arbeit

Im weiteren Verlauf der Arbeit werden zunächst die theoretischen Grundlagen vorgestellt, gefolgt von der Beschreibung des Datensatzes, der Methodik und der Ergebnisse. Abschliessend werden die Erkenntnisse diskutiert und potenzielle Limitationen sowie Ansätze für zukünftige Forschung aufgezeigt.

# Theoretische Grundlagen

## Kreditrisiko und Zahlungsausfälle

Das Kreditrisiko bezeichnet die Gefahr, dass ein Kreditnehmer seinen vertraglichen Zahlungsverpflichtungen gegenüber dem Kreditgeber nicht oder nur unvollständig nachkommt. Dies kann zu finanziellen Verlusten für den Kreditgeber führen. Das Kreditrisiko ist somit eine zentrale Komponente im Risikomanagement von Banken und Finanzinstituten.[[1]](#footnote-1)

Ein **Zahlungsausfall** tritt ein, wenn der Kreditnehmer nicht in der Lage ist, die vereinbarten Zins- und Tilgungszahlungen fristgerecht zu leisten. Solche Ausfälle können verschiedene Ursachen haben, darunter finanzielle Schwierigkeiten des Kreditnehmers, wirtschaftliche Abschwünge oder unerwartete Ereignisse wie Arbeitslosigkeit oder Krankheit. Die Folgen von Zahlungsausfällen sind für Kreditgeber gravierend, da sie direkte finanzielle Verluste bedeuten und die Liquidität sowie die Kapitalausstattung der Institution beeinträchtigen können.

Um das Kreditrisiko zu bewerten und Zahlungsausfälle zu prognostizieren, setzen Finanzinstitute auf verschiedene Methoden und Modelle. Traditionell wurden statistische Ansätze wie die Logistische Regression verwendet, um die Wahrscheinlichkeit eines Zahlungsausfalls basierend auf historischen Daten und bestimmten Kreditnehmermerkmalen zu schätzen. Mit dem Fortschritt in der Datenverarbeitung und der Verfügbarkeit grosser Datenmengen gewinnen jedoch zunehmend komplexere Machine-Learning-Modelle an Bedeutung. Modelle wie der **Random Forest** oder **XGBoost** können nichtlineare Zusammenhänge in den Daten erfassen und bieten somit potenziell präzisere Vorhersagen. Allerdings erfordern sie auch eine sorgfältige Validierung und Kalibrierung, um Überanpassungen zu vermeiden und die Generalisierbarkeit sicherzustellen.

Die effektive Bewertung des Kreditrisikos und die Vorhersage von Zahlungsausfällen sind essenziell, um Kreditentscheidungen fundiert zu treffen, Zinssätze angemessen zu gestalten und regulatorische Anforderungen zu erfüllen. Ein präzises Risikomanagement trägt dazu bei, die Stabilität des Finanzsystems zu gewährleisten und das Vertrauen der Anleger und Kunden in die Finanzmärkte aufrechtzuerhalten.

## Klassifikationsprobleme im maschinellen Lernen

Im Bereich des maschinellen Lernens stellen **Klassifikationsprobleme** eine zentrale Aufgabe dar. Sie zielen darauf ab, Dateninstanzen einer oder mehreren vordefinierten Kategorien zuzuordnen. Dies geschieht anhand von Input-Daten und deren Merkmalen, wobei ein Klassifikationsmodell anhand eines Trainingsdatensatzes erstellt wird, der sowohl die Merkmale als auch die zugehörigen Kategorien (Labels) enthält.

### Definition und Grundlagen

Ein Klassifikationsproblem liegt vor, wenn das Ziel darin besteht, eine qualitative Variable (z. B. „ja/nein“, „Kategorie A/B/C“) vorherzusagen. Typische Beispiele sind:

* Vorhersage, ob ein Kunde einen Kredit zurückzahlt (Zahlungsausfall: ja/nein).
* Identifikation von Spam in E-Mails (Spam/Nicht-Spam).
* Klassifikation von medizinischen Diagnosen basierend auf Patientendaten.

Klassifikationsprobleme gehören zur **überwachten Lernmethode** im maschinellen Lernen. Dabei wird ein Modell erstellt, das die Beziehung zwischen den Eingabedaten (Features) und den Zielvariablen (Labels) erlernt.

### Algorithmische Ansätze

Es existieren zahlreiche Algorithmen, um Klassifikationsprobleme zu lösen. Diese unterscheiden sich hinsichtlich ihrer Komplexität, Leistung und den zugrunde liegenden Annahmen über die Daten. Zu den bekanntesten Algorithmen zählen:

1. **Logistische Regression:**  
   Ein einfacher, aber effektiver Ansatz, insbesondere für binäre Klassifikationsprobleme. Die logistische Regression modelliert die Wahrscheinlichkeit einer Kategorie anhand einer Sigmoid-Funktion und ist insbesondere dann geeignet, wenn die Datenlinear separierbar sind.
2. **Entscheidungsbäume und Random Forests**:  
   Entscheidungsbäume basieren auf einer hierarchischen Struktur, die Entscheidungen durch logische Bedingungen trifft. Random Forests kombinieren mehrere Entscheidungsbäume (Ensemble-Ansatz), um die Vorhersagegenauigkeit zu erhöhen und Überanpassungen zu reduzieren.
3. **Gradient Boosting-Algorithmen (z. B. XGBoost)**:  
   Diese Ansätze verbessern die Leistung eines schwachen Lerners (z. B. eines einfachen Entscheidungsbaums) durch iterative Fehlerreduktion. Gradient Boosting ist besonders leistungsfähig bei grossen und komplexen Datensätzen.
4. **Neuronale Netze**:  
   Diese Modelle sind inspiriert von biologischen Gehirnstrukturen und können komplexe, nichtlineare Beziehungen in den Daten modellieren. Sie eignen sich besonders für grosse Datensätze mit vielen Features.

### Leistungsbewertung von Klassifikationsmodellen

Die Bewertung der Güte eines Klassifikationsmodells ist entscheidend, um dessen Praxistauglichkeit sicherzustellen. Häufig eingesetzte Metriken sind:

* **Genauigkeit (Accuracy)**: Anteil korrekt klassifizierter Instanzen.
* **Präzision und Recall**: Fokus auf die Vorhersagegenauigkeit und -vollständigkeit einzelner Klassen.
* **F1-Score**: Harmonisches Mittel aus Präzision und Recall, besonders bei unausgewogenen Datensätzen relevant.
* **ROC-AUC (Receiver Operating Characteristic - Area Under Curve)**: Mass für die Trennfähigkeit des Modells über verschiedene Schwellenwerte hinweg.

### Herausforderungen und Limitationen

Trotz ihrer breiten Anwendbarkeit können Klassifikationsprobleme mit Herausforderungen verbunden sein:

* **Datenqualität**: Unvollständige oder fehlerhafte Daten beeinflussen die Modellleistung.
* **Unausgewogene Datensätze**: Wenn eine Klasse stark überrepräsentiert ist, kann dies zu Verzerrungen führen.
* **Interpretierbarkeit**: Komplexe Modelle wie XGBoost oder neuronale Netze sind oft schwer verständlich, was bei Entscheidungen im Finanzwesen problematisch sein kann.

Die Anwendung von Klassifikationsmodellen in der Kreditkarten-Ausfallanalyse erfordert eine sorgfältige Auswahl und Evaluierung geeigneter Algorithmen, um robuste und verlässliche Vorhersagen zu ermöglichen.

## Überblick über die untersuchten Machine-Learning-Modelle

In diesem Abschnitt werden die Machine-Learning-Modelle vorgestellt, die in der vorliegenden Arbeit zur Vorhersage von Zahlungsausfällen bei Kreditkartenkunden eingesetzt werden. Der Fokus liegt auf drei häufig verwendeten und leistungsfähigen Ansätzen: der logistischen Regression, dem Random Forest und dem XGBoost-Algorithmus. Diese Modelle wurden aufgrund ihrer unterschiedlichen Eigenschaften und ihrer etablierten Anwendung in Klassifikationsproblemen ausgewählt.

### Logistische Regression

Die **logistische Regression** ist ein einfaches und interpretierbares Modell für binäre Klassifikationsprobleme. Sie modelliert die Wahrscheinlichkeit, dass eine Instanz zu einer bestimmten Klasse gehört, mithilfe einer Sigmoid-Funktion:

* **Stärken**:
  + Einfache Implementierung und schnelle Berechnung.
  + Hohe Interpretierbarkeit der Koeffizienten, was insbesondere in regulierten Bereichen wie dem Finanzwesen von Vorteil ist.
  + Geeignet für lineare Entscheidungsgrenzen.
* **Schwächen**:
  + Begrenzte Leistungsfähigkeit bei nichtlinearen Beziehungen zwischen Features und Zielvariablen.
  + Anfällig für Multikollinearität zwischen den Features.

Die logistische Regression ist oft ein Ausgangspunkt für die Analyse, da sie grundlegende Einblicke in die zugrunde liegende Datenstruktur bietet.

### Random Forest

Der **Random Forest** ist ein Ensemble-Lernverfahren, das auf der Aggregation von Entscheidungsbäumen basiert. Jeder Baum wird auf einem zufälligen Teil des Trainingsdatensatzes und einer zufälligen Untermenge der Features trainiert. Die finale Vorhersage erfolgt durch Mehrheitsvotum (Klassifikation) oder Mittelwertbildung (Regression).

* **Stärken**:
  + Robust gegenüber Überanpassung durch die Kombination mehrerer Bäume.
  + Kann komplexe, nichtlineare Beziehungen modellieren.
  + Handhabung von unausgewogenen Datensätzen durch Gewichtung der Klassen möglich.
* **Schwächen**:
  + Weniger interpretierbar als ein einzelner Entscheidungsbaum.
  + Rechenintensiv bei sehr grossen Datensätzen.

Random Forests sind bekannt für ihre Vielseitigkeit und Stabilität und eignen sich gut für Probleme, bei denen die Beziehung zwischen den Variablen nichtlinear und komplex ist.

### – XGBoost

Der **Extreme Gradient Boosting (XGBoost)**-Algorithmus ist eine Erweiterung des Gradient-Boosting-Ansatzes, der Entscheidungsbäume iterativ trainiert, um Fehler der vorherigen Bäume zu reduzieren. XGBoost ist für seine Effizienz, Genauigkeit und Anpassungsfähigkeit bekannt.

* **Stärken**:
  + Hervorragende Leistung bei strukturierten Daten.
  + Umfassende Funktionen zur Regularisierung, die Überanpassung minimieren.
  + Hohe Flexibilität bei der Optimierung durch anpassbare Verlustfunktionen.
* **Schwächen**:
  + Erhöhte Komplexität im Vergleich zu anderen Algorithmen.
  + Geringere Interpretierbarkeit als einfache Modelle wie die logistische Regression.

XGBoost hat sich als leistungsstarker Algorithmus in vielen Machine-Learning-Wettbewerben bewährt und ist besonders geeignet für grosse Datensätze mit vielen Features und komplexen Mustern.

# Datensatzbeschreibung und Explorative Datenanalyse

## Herkunft und Struktur des Datensatzes

Der verwendete Datensatz, bekannt als "Default of Credit Card Clients Dataset", stammt aus Taiwan und umfasst Informationen zu Kreditkartenkunden von April 2005 bis September 2005. Der Datensatz wurde ursprünglich im UCI Machine Learning Repository bereitgestellt und beinhaltet Daten zu Demografie, Kreditverhalten, Zahlungshistorie und Rechnungsinformationen.

## Beschreibung der Variablen

Der Datensatz besteht aus 25 Variablen, die im Folgenden beschrieben werden:

* **ID**: Eindeutige Kennung jedes Kunden.
* **LIMIT\_BAL**: Kreditlimit in NT-Dollar (einschliesslich individueller und familiärer Kredite).
* **SEX**: Geschlecht (1 = männlich, 2 = weiblich).
* **EDUCATION**: Bildungsniveau (1 = Graduiertenschule, 2 = Universität, 3 = Gymnasium, 4 = andere, 5 und 6 = unbekannt).
* **MARRIAGE**: Familienstand (1 = verheiratet, 2 = ledig, 3 = andere).
* **AGE**: Alter in Jahren.
* **PAY\_0 bis PAY\_6**: Rückzahlungsstatus von April 2005 bis September 2005 (-1 = pünktliche Zahlung, 1 = Zahlungsverzug für einen Monat, bis 8 = acht Monate Zahlungsverzug).
* **BILL\_AMT1 bis BILL\_AMT6**: Rechnungsbeträge von April 2005 bis September 2005 (in NT-Dollar).
* **PAY\_AMT1 bis PAY\_AMT6**: Beträge der getätigten Zahlungen von April 2005 bis September 2005 (in NT-Dollar).
* **default.payment.next.month**: Zielvariable, die angibt, ob eine Zahlungsausfälle im nächsten Monat auftrat (1 = ja, 0 = nein).

Der Datensatz umfasst insgesamt 30.000 Beobachtungen.

## Explorative Datenanalyse (EDA)

### Ziel der Analyse

Die explorative Datenanalyse hat das Ziel, ein grundlegendes Verständnis über die Datenstruktur, Verteilungen und Beziehungen zwischen den Variablen zu gewinnen. Dabei werden sowohl deskriptive Statistiken als auch grafische Methoden verwendet.

### Verteilung der Zielvariablen

Die Zielvariable **default.payment.next.month** ist binär und zeigt die Kreditkartenrückzahlungsbereitschaft der Kunden an. Ein erster Blick auf die Verteilung zeigt:

* **Nicht-Ausfälle (0)**: 23.364 (77,88 %)
* **Ausfälle (1)**: 6.636 (22,12 %)

### Analyse der numerischen Variablen

Die numerischen Variablen wie **LIMIT\_BAL**, **AGE**, **BILL\_AMT1-6** und **PAY\_AMT1-6** wurden auf Verteilungen und Ausreisser untersucht. Die Ergebnisse zeigen:

* **LIMIT\_BAL**: Das Kreditlimit variiert zwischen 10.000 und 1.000.000 NT-Dollar, wobei der Median bei 140.000 NT-Dollar liegt.
* **AGE**: Das Alter der Kunden reicht von 21 bis 79 Jahren, mit einem Durchschnitt von 35 Jahren.

### Analyse der kategorischen Variablen

Die Verteilungen der kategorischen Variablen zeigen:

* **SEX**: 60 % der Kunden sind weiblich, 40 % männlich.
* **EDUCATION**: Der Grossteil der Kunden hat einen Universitätsabschluss (47 %).
* **MARRIAGE**: 53 % der Kunden sind ledig, 46 % verheiratet, und 1 % befinden sich in einer anderen Beziehungsform.

### Korrelation zwischen Variablen

Die Korrelation zwischen den numerischen Variablen wurde analysiert, um potenzielle Zusammenhänge zu identifizieren. Hierbei zeigen **BILL\_AMT1-6** und **PAY\_AMT1-6** eine moderate Korrelation, was auf ähnliche Zahlungstrends hinweist.

# Methodik

## Datenaufbereitung und Feature Engineering

### Ausreisseranalyse

Die Analyse von Ausreissern ist ein entscheidender Schritt in der Datenvorverarbeitung, da Ausreisser das Verhalten von Modellen erheblich beeinflussen können. In dieser Sektion werden die Ergebnisse der Ausreisseranalyse für die Variablen der Kategorien "BILL\_AMT" und "PAY\_AMT" präsentiert, basierend auf der IQR-Methode und der Z-Score-Methode.

#### IQR-Methode

Die Interquartilsabstand-Methode (IQR-Methode) identifiziert Ausreiser als Werte, die auserhalb der unteren und oberen Grenzen liegen, welche wie folgt berechnet werden:[[2]](#footnote-3)

##### Ergebnisse der IQR-Analyse für "BILL\_AMT":

* **BILL\_AMT1:** 2400 Ausreisser (IQR: 63.532,25; Grenzen: [-91.739,62, 162.389,38])
* **BILL\_AMT2:** 2395 Ausreisser (IQR: 61.021,50; Grenzen: [-88.547,50, 155.538,50])
* **BILL\_AMT3:** 2469 Ausreisser (IQR: 57.498,50; Grenzen: [-83.581,50, 146.412,50])
* **BILL\_AMT4:** 2622 Ausreisser (IQR: 52.179,25; Grenzen: [-75.942,12, 132.774,88])
* **BILL\_AMT5:** 2725 Ausreisser (IQR: 48.427,50; Grenzen: [-70.878,25, 122.831,75])
* **BILL\_AMT6:** 2693 Ausreisser (IQR: 47.942,25; Grenzen: [-70.657,38, 121.111,62])

##### Ergebnisse der IQR-Analyse für "PAY\_AMT":

* **PAY\_AMT1:** 2745 Ausreisser (IQR: 4.006,00; Grenzen: [-5.009,00, 11.015,00])
* **PAY\_AMT2:** 2714 Ausreisser (IQR: 4.167,00; Grenzen: [-5.417,50, 11.250,50])
* **PAY\_AMT3:** 2598 Ausreisser (IQR: 4.115,00; Grenzen: [-5.782,50, 10.677,50])
* **PAY\_AMT4:** 2994 Ausreisser (IQR: 3.717,25; Grenzen: [-5.279,88, 9.589,12])
* **PAY\_AMT5:** 2945 Ausreisser (IQR: 3.779,00; Grenzen: [-5.416,00, 9.700,00])
* **PAY\_AMT6:** 2958 Ausreisser (IQR: 3.882,25; Grenzen: [-5.705,62, 9.823,38])

#### Z-Score-Methode

Die Z-Score-Methode identifiziert Ausreisser als Werte, deren Abstand vom Mittelwert mehr als eine festgelegte Anzahl von Standardabweichungen (Schwellenwert) beträgt. Der Z-Score wird wie folgt berechnet:[[3]](#footnote-4)

##### Ergebnisse der Z-Score-Analyse für "BILL\_AMT":

* **BILL\_AMT1:** 686 Ausreisser (Schwellenwert: 3)
* **BILL\_AMT2:** 670 Ausreisser (Schwellenwert: 3)
* **BILL\_AMT3:** 661 Ausreisser (Schwellenwert: 3)
* **BILL\_AMT4:** 680 Ausreisser (Schwellenwert: 3)
* **BILL\_AMT5:** 651 Ausreisser (Schwellenwert: 3)
* **BILL\_AMT6:** 651 Ausreisser (Schwellenwert: 3)

#### Visualisierung der Ergebnisse

Um die Verteilung der Werte und die identifizierten Ausreisser besser zu verstehen, wurden Histogramme erstellt. Diese zeigen die originalen Verteilungen sowie die Verteilungen nach Transformationen, wie z. B. logarithmischen Transformationen.

**Beispiele:**

* Die Abbildung "Original Verteilung von BILL\_AMT1 mit Z-Score-Schwelle (3)" zeigt die identifizierten Ausreisser als rote Linie
* Die logarithmische Transformation verbessert die Visualisierung, indem extreme Werte besser skaliert dargestellt werden.

A graph of a line

Description automatically generated with medium confidence

* Die Boxplot-Diagramme, die mit der IQR-Methode erstellt wurden, zeigen die identifizierten Ausreisser als Punkte ausserhalb der unteren und oberen Grenzen (Whiskers). Diese Visualisierungen verdeutlichen, wie stark einzelne Werte von der zentralen Verteilung abweichen.
* Im Boxplot-Diagramm ist erkennbar, dass die Dichte der Daten nach rechts hin kontinuierlich abnimmt, was mit den Ergebnissen der Z-Score-Analyse übereinstimmt.

A graph of a box plot

Description automatically generated

### Entscheidung zur Ausreisserbehandlung

Nach sorgfältiger Analyse und Abwägung wurde entschieden, die Z-Score-Methode für die Identifikation und Entfernung von Ausreissern zu verwenden. Der Hauptgrund dafür liegt in der Anzahl der identifizierten Ausreisser:

* Die IQR-Methode markiert fast 10 % der Werte je Spalte (etwa 2900 pro Spalte) als Ausreisser, was einen erheblichen Einfluss auf den Datensatz hat.
* Die Z-Score-Methode identifiziert hingegen eine deutlich geringere Anzahl von Ausreissern (etwa 600 pro Spalte), was weniger Eingriffe in die Datenanalyse erfordert.

Ein weiterer Vorteil der Z-Score-Methode ist ihre Strenge in der Definition extremer Werte, wodurch sie sich besser für diesen Datensatz eignet. Indem nur die extremsten Werte entfernt werden, bleibt die Mehrheit der Daten unberührt, und der Informationsgehalt des Datensatzes bleibt weitgehend erhalten.

Das Entfernen der Ausreisser war zwar für XGBoost und Random Forest technisch nicht notwendig, hat aber keine negativen Auswirkungen, da diese Modelle robust gegenüber Ausreissern sind. Dennoch wurde diese Entscheidung getroffen, da sie für die logistische Regression potenziell hilfreich ist. Das Entfernen der Ausreisser könnte hier die Stabilität der Berechnungen verbessern, insbesondere bei der Schätzung von Modellparametern, die durch extreme Werte verfälscht werden könnten.

Darüber hinaus trägt das Entfernen der Ausreisser zur allgemeinen Robustheit der Analysen bei. Selbst wenn Ausreisser für XGBoost und Random Forest weniger kritisch sind, schafft die Bereinigung der Daten eine einheitlichere Basis für den Vergleich der Modellleistungen. Dies ermöglicht es, die Auswirkungen von Modellentscheidungen isolierter zu bewerten.

In den nächsten Schritten wird die Datenanalyse mit den bereinigten Daten fortgesetzt, um zu untersuchen, ob die Entfernung der Ausreisser das Modellverhalten verbessert. Gleichzeitig wird geprüft, ob die entfernten Werte tatsächlich störend oder potenziell informativ waren. Diese iterative Vorgehensweise stellt sicher, dass die Datenqualität verbessert wird, ohne relevante Informationen zu verlieren.[[4]](#footnote-5)

### Datenumwandlung

Die Datenumwandlung ist ein wichtiger Schritt in der Datenaufbereitung, um sicherzustellen, dass die Modelle effizient und korrekt arbeiten können. In diesem Kapitel werden die angewandten Methoden der Datenumwandlung beschrieben.

#### Feature-Skalierung

Die Skalierung der numerischen Variablen wurde mit dem MinMax Scaler durchgeführt. Diese Methode transformiert die Werte so, dass sie sich in einem Bereich von 0 bis 1 befinden, wobei die Proportionen der Ursprungsdaten erhalten bleiben. Dies erleichtert den Machine-Learning-Modellen die Verarbeitung der Daten, da sie empfindlich auf unterschiedliche Skalen reagieren können.

Die folgenden Spalten wurden skaliert:

* **LIMIT\_BAL** (Kreditlimit)
* **AGE** (Alter)
* **BILL\_AMT**\* (Rechnungsbeträge der letzten sechs Monate)
* **PAY\_AMT**\* (Zahlungsbeträge der letzten sechs Monate)

Nach der Skalierung sind alle Werte dieser Spalten im Bereich von 0 bis 1, wodurch eine bessere Vergleichbarkeit zwischen den Features gewährleistet wird.[[5]](#footnote-6)

#### One Hot Encoding

Die Verarbeitung der kategorialen Variablen erfolgte nach dem One-Hot-Encoding-Verfahren. Dabei wurden die Originalspalten durch neue, binäre Spalten ersetzt, um den Modellen die Interpretation zu erleichtern. Die folgenden Variablen wurden bearbeitet:

* **SEX** (Geschlecht)
* **EDUCATION** (Bildungsstand)
* **MARRIAGE** (Familienstand)

Die ursprünglichen Spalten wurden gelöscht und durch die folgenden neuen Spalten ersetzt:

* **MALE** (1 = männlich, 0 = weiblich)
* **FEMALE** (1 = weiblich, 0 = männlich)
* **GRADUATE\_SCHOOL** (1 = Hochschulabschluss, 0 = andere)
* **UNIVERSITY** (1 = Universitätsabschluss, 0 = andere)
* **HIGH\_SCHOOL** (1 = Abitur, 0 = andere)
* **OTHERS\_EDUCATION** (1 = andere Bildung, 0 = übrige Kategorien)
* **UNKNOWN\_EDUCATION** (1 = unbekannter Bildungsstand, 0 = andere)
* **MARRIED** (1 = verheiratet, 0 = andere)
* **SINGLE** (1 = ledig, 0 = andere)
* **OTHER\_RELATIONSHIP** (1 = andere Beziehung, 0 = übrige Kategorien)

Durch dieses Verfahren wurde sichergestellt, dass die Modelle keine falschen Annahmen über die Reihenfolge oder den Abstand zwischen den Kategorien treffen. Die binären Werte (0 oder 1) erleichtern den Modellen das Lernen und verbessern die Vorhersagegenauigkeit.[[6]](#footnote-7)

### Feature Engineering

Im Rahmen der Datenaufbereitung wurden zusätzliche Merkmale (Features) generiert, um potenzielle Zusammenhänge zwischen bestehenden Variablen zu identifizieren und die Leistung der Machine-Learning-Modelle zu verbessern. Dabei wurden insbesondere Verhältnisse wie **BILL\_AMT / LIMIT\_BAL** und **PAY\_AMT / BILL\_AMT** eingeführt. Diese neuen Features liefern wertvolle Informationen, die in den ursprünglichen Daten nicht explizit enthalten sind.

#### Die Wahl der Verhältnisse:

**BILL\_AMT / LIMIT\_BAL**

* Dieses Verhältnis gibt an, wie stark das Kreditlimit (LIMIT\_BAL) durch die aktuelle Rechnung (BILL\_AMT) ausgeschöpft ist. Es ist ein Indikator für das Verbraucherverhalten und das Risiko.
  + **Höhere Werte** (z. B. nahe 1) deuten darauf hin, dass ein Kunde sein Kreditlimit fast voll ausschöpft, was auf ein potenziell höheres Ausfallrisiko hinweisen könnte.
  + **Niedrigere Werte** deuten darauf hin, dass der Kunde sein Kreditlimit nur geringfügig nutzt, was auf eine konservativere Nutzung hindeuten könnte.
* Dieses Verhältnis ermöglicht es den Modellen, die Kreditnutzung eines Kunden in Relation zum verfügbaren Limit besser einzuschätzen.

**PAY\_AMT / BILL\_AMT**

* Dieses Verhältnis beschreibt, wie viel ein Kunde im Vergleich zu seiner Rechnung (BILL\_AMT) zurückgezahlt hat. Es zeigt das Zahlungsverhalten und die finanzielle Disziplin.
  + **Höhere Werte** (z. B. über 1) deuten darauf hin, dass der Kunde mehr als seine offene Rechnung bezahlt hat, was positiv für die Kreditwürdigkeit sein könnte.
  + **Werte nahe 0** zeigen, dass der Kunde entweder nur minimale oder keine Zahlungen leistet, was ein Risiko für zukünftige Ausfälle darstellen könnte.
* Dieses Verhältnis hilft, Kunden mit guten Rückzahlungspraktiken von potenziell riskanteren Kunden zu unterscheiden.

#### Nutzen der Verhältnisse

Die Einführung solcher Verhältnisse verbessert die Übersichtlichkeit der Daten für die Algorithmen und unterstützt sie dabei, bessere Muster in den Daten zu erkennen. Durch diese relativen Merkmale können die Modelle kontextuelle Zusammenhänge verstehen, die bei der Verwendung der rohen Features (z. B. BILL\_AMT oder LIMIT\_BAL allein) möglicherweise übersehen worden wären.

Zusätzlich können diese Verhältnisse auch dabei helfen, das Modellverhalten zu interpretieren, da sie direkt auf finanzielle Praktiken und Risiken hinweisen. Dies macht sie nicht nur für prädiktive Zwecke nützlich, sondern auch für die Erklärung der Modellentscheidungen.

## Implementierung der Machine-Learning-Modelle

### Performancemetriken zur Bewertung der Modelle

Die Bewertung der Modelle in diesem Projekt basiert auf einer Vielzahl von Performancemetriken, die unterschiedlichen Aspekte der Modellleistung beleuchten. Jede dieser Metriken bietet einen einzigartigen Blickwinkel auf die Genauigkeit und Zuverlässigkeit der Vorhersagen. Nachfolgend werden die wichtigsten Metriken und ihre Bedeutung erläutert.

#### Accuracy, Precision, Recall, F1-Score

a Accuracy (Genauigkeit): Die Accuracy misst den Anteil der korrekt klassifizierten Datenpunkte an der Gesamtanzahl der Datenpunkte. Sie wird folgendermassen berechnet:

Dabei stehen:

* **TP (True Positives):** Korrekt als positiv klassifizierte Fälle.
* **TN (True Negatives):** Korrekt als negativ klassifizierte Fälle.
* **FP (False Positives):** Fälschlich als positiv klassifizierte Fälle.
* **FN (False Negatives):** Fälschlich als negativ klassifizierte Fälle.

**Bedeutung:** Die Accuracy bietet eine allgemeine Bewertung der Modellleistung. Allerdings kann sie bei unausgewogenen Datensätzen irreführend sein, da sie dominante Klassen bevorzugt.

**Precision (Präzision):** Die Precision zeigt, wie viele der als positiv vorhergesagten Fälle tatsächlich positiv sind:

**Bedeutung:** Precision ist besonders wichtig, wenn die Kosten für falsch-positive Ergebnisse hoch sind, z. B. bei der Vergabe von Krediten an nicht kreditwürdige Kunden.

**Recall (Sensitivität):** Recall misst, wie viele der tatsächlich positiven Fälle korrekt erkannt wurden:

**Bedeutung:** Recall ist entscheidend, wenn das Ziel darin besteht, möglichst viele positive Fälle zu identifizieren, z. B. zur Erkennung von Zahlungsausfällen.

**F1-Score:** Der F1-Score ist das harmonische Mittel von Precision und Recall und bietet eine ausgewogene Bewertung der Modellleistung, insbesondere bei unausgewogenen Datensätzen:

**Bedeutung:** Der F1-Score ist hilfreich, um ein Gleichgewicht zwischen Precision und Recall zu bewerten. Ein hoher F1-Score zeigt, dass das Modell sowohl eine hohe Präzision als auch einen hohen Recall aufweist.

**Zusammenfassung:** Die Kombination dieser Metriken liefert ein umfassendes Bild der Modellleistung. Während die Accuracy einen allgemeinen Überblick bietet, erlauben Precision und Recall eine gezieltere Analyse der Modellgenauigkeit für spezifische Klassen. Der F1-Score stellt sicher, dass weder Precision noch Recall vernachlässigt werden. Diese Metriken sind besonders relevant für Kreditunternehmen, da sie die Balance zwischen der Minimierung von Risiken (z. B. falsche Kreditzusagen) und der Maximierung von Chancen (z. B. korrekte Identifikation kreditwürdiger Kunden) sicherstellen.

#### ROC-Kurve und AUC-Wert

Die ROC-Kurve (Receiver Operating Characteristic) und der AUC-Wert (Area Under the Curve) sind zentrale Werkzeuge zur Bewertung der Modellleistung, insbesondere bei Klassifikationsproblemen mit zwei Klassen. Diese Metriken helfen, die Trennfähigkeit eines Modells zu quantifizieren und die Balance zwischen den Klassifikationsentscheidungen (z. B. Zahlungsausfälle vs. keine Zahlungsausfälle) besser zu verstehen.

##### ROC-Kurve

Die ROC-Kurve stellt die Sensitivität (Recall) gegen die 1-Spezifität (False Positive Rate) für verschiedene Schwellenwerte dar. Sie zeigt auf, wie sich das Verhältnis von True Positives zu False Positives mit der Veränderung des Schwellenwerts für die Klassifikation entwickelt.

* **True Positive Rate (TPR / Recall):** Anteil der korrekt erkannten positiven Fälle an allen tatsächlichen positiven Fällen:
* **False Positive Rate (FPR):** Anteil der fälschlich als positiv klassifizierten Fälle an allen tatsächlichen negativen Fällen:

Bedeutung:  
Eine ideale ROC-Kurve verläuft nahe an der oberen linken Ecke, was auf eine hohe Sensitivität (Recall) und gleichzeitig eine niedrige False Positive Rate hinweist. Modelle, die zufällig zwischen den Klassen unterscheiden, weisen hingegen eine diagonale ROC-Kurve auf.

##### AUC-Wert (Area Under the Curve)

Der AUC-Wert gibt die Fläche unter der ROC-Kurve an und bietet eine quantitative Bewertung der Trennschärfe eines Modells. Der Wert liegt zwischen 0 und 1:

* **AUC = 1.0:** Perfekte Trennschärfe (alle Klassen korrekt unterschieden).
* **AUC = 0.5:** Keine Trennschärfe (Modell rät zufällig).
* **AUC < 0.5:** Schlechter als zufälliges Raten (Modell klassifiziert systematisch falsch).

Bedeutung:  
Ein höherer AUC-Wert zeigt, dass das Modell die Klassen besser unterscheiden kann. In diesem Projekt wurde ein AUC-Wert von **0.7870** erzielt, was auf eine gute, jedoch nicht perfekte Trennfähigkeit hinweist.

##### Vorteile der ROC-Kurve und des AUC-Werts

1. **Unabhängigkeit vom Schwellenwert:**  
   Im Gegensatz zu Metriken wie Accuracy oder Precision, die von einem festen Schwellenwert abhängen, berücksichtigt die ROC-Kurve die Modellleistung über alle möglichen Schwellenwerte hinweg. Dadurch wird ein umfassenderes Bild der Modellqualität ermöglicht.
2. **Vergleichbarkeit von Modellen:**  
   Die AUC erlaubt einen direkten Vergleich zwischen verschiedenen Modellen. Ein Modell mit höherem AUC-Wert wird in der Regel bevorzugt, da es eine bessere Trennschärfe zwischen den Klassen aufweist.
3. **Praktische Relevanz für Kreditunternehmen:**  
   Ein hohes AUC zeigt, dass das Modell Zahlungsausfälle (Klasse 1) gut von erfolgreichen Rückzahlungen (Klasse 0) unterscheiden kann, unabhängig vom Schwellenwert. Dies ist besonders nützlich, wenn das Unternehmen flexibel Schwellenwerte anpassen möchte, um z. B. die Anzahl der False Positives oder False Negatives je nach Geschäftsstrategie zu minimieren.[[7]](#footnote-8)

#### Zusammenfassung

Die ROC-Kurve bietet eine visuelle Darstellung der Modellleistung über verschiedene Schwellenwerte hinweg, während der AUC-Wert eine numerische Zusammenfassung der Trennschärfe liefert. Im Kontext dieses Projekts ermöglicht der AUC-Wert von beispielsweise **0.7870**, wie bei XGBoost, ein zuverlässiges Unterscheiden zwischen Kreditrückzahlungen und Zahlungsausfällen, wobei noch Raum für Optimierungen besteht. Für das Kreditunternehmen ist die ROC-Kurve ein wertvolles Werkzeug, um Schwellenwerte anzupassen und Risiken sowie Chancen dynamisch zu steuern.

### XGBoost

#### Cross-Validation Scores

* **Cross-Validation Scores:** [0.81017882, 0.8131591, 0.82003668, 0.8149507, 0.81563861]
* **Mean CV Score:** 0.8148 (+/- 0.0065)

**Bedeutung:** Die Cross-Validation-Scores zeigen die Konsistenz des Modells über verschiedene Datensplits hinweg. Ein Mean CV Score von 81.48 % bedeutet, dass das Modell in etwa 81.5 % der Fälle die Zielvariable korrekt vorhersagt. Die geringe Standardabweichung (+/- 0.0065) deutet auf eine stabile Leistung hin, unabhängig vom Trainingsdatensplit.

#### Test Accuracy

* **Test Accuracy:** 0.8150

**Bedeutung:** Das Modell hat auf den Testdaten eine Genauigkeit von 81.5 %, was insgesamt positiv ist. Allerdings gibt die Genauigkeit allein wenig Aufschluss über das Gleichgewicht zwischen den Klassen, insbesondere bei stark unausgeglichenen Datensätzen.

#### Classification Report

|  |  |  |
| --- | --- | --- |
| Metric | Klasse 0 (Nicht-Default) | Klasse 1 (Default) |
| Precision | 0.84 | 0.67 |
| Recall | 0.95 | 0.37 |
| F1-Score | 0.89 | 0.48 |

**Bedeutung der Metriken:**

* **Precision (Genauigkeit):**
  + Bei Klasse 1 (Default) liegt die Präzision bei 67 %, was bedeutet, dass 67 % der als Default klassifizierten Fälle korrekt sind.
* **Recall (Sensitivität):**
  + Klasse 1 hat nur einen Recall von 37 %, was bedeutet, dass das Modell Schwierigkeiten hat, tatsächliche Default-Fälle zu erkennen.
* **F1-Score:**
  + Ein kombinierter Wert aus Precision und Recall, der zeigt, dass die Modellleistung für Klasse 1 verbesserungswürdig ist.

**Fazit:** Das Modell funktioniert gut für Klasse 0 (Nicht-Default), hat jedoch Probleme, die weniger häufige Klasse 1 (Default) zu identifizieren. Dies liegt wahrscheinlich an einem Klassenungleichgewicht im Datensatz.

#### Anpassung der Klassengewichte

Da die Zielvariable ungleich verteilt ist, wurde eine Klassengewichtung (“Class Weights”) eingeführt, um die Minderheitsklasse (Klasse 1) höher zu gewichten und ein besseres Gleichgewicht zwischen den Klassen zu erzielen. Der Parameter scale\_pos\_weight wurde wie folgt berechnet:

Durch diese Gewichtung werden Fehler bei der Klassifizierung von Defaults (Klasse 1) stärker bestraft. Es wurde bewusst auf Oversampling und Undersampling verzichtet, um sicherzustellen, dass alle Modelle mit denselben Datensätzen arbeiten.

##### Ergebnisse nach Anpassung der Gewichtung

**Cross-Validation Scores**

* **Cross-Validation Scores:** [0.75263641, 0.75080238, 0.76593306, 0.75693648, 0.75510204]
* **Mean CV Score:** 0.7563 (+/- 0.0105)

**Test Accuracy**

* **Test Accuracy:** 0.7647

**Classification Report**

|  |  |  |
| --- | --- | --- |
| Metric | Klasse 0 (Nicht-Default) | Klasse 1 (Default) |
| Precision | 0.88 | 0.49 |
| Recall | 0.80 | 0.63 |
| F1-Score | 0.84 | 0.55 |

**Bedeutung:** Nach der Anpassung hat sich der **Recall** der Klasse 1 (Default) von 37 % auf 63 % erheblich verbessert. Dies zeigt, dass das Modell nun deutlich besser darin ist, tatsächliche Defaults zu erkennen. Die Precision der Klasse 1 ist jedoch leicht gesunken, was auf eine Zunahme von False Positives hinweist. Insgesamt ist die Balance zwischen den Klassen jedoch verbessert.

**Feature-Importance**

Die wichtigsten Merkmale laut XGBoost:

|  |  |
| --- | --- |
| Feature | Bedeutung (%) |
| PAY\_0 (Zahlungsstatus) | 30.85 |
| PAY\_2 (Zahlungsverzögerung) | 5.56 |
| PAY\_4 | 4.59 |
| PAY\_3 | 3.58 |
| PAY\_5 | 3.51 |
| PAY\_6 | 3.49 |
| PAY\_AMT3 (Zahlung 3. Monat) | 2.34 |
| CREDIT\_UTILIZATION\_RATIO\_3 | 1.98 |
| LIMIT\_BAL (Kreditlimit) | 1.90 |
| PAY\_AMT2 (Zahlung 2. Monat) | 1.87 |

**Bedeutung:**

* PAY\_0 ist das wichtigste Merkmal mit 30.85 % Einfluss. Dies ist erwartungsgemäss, da der Zahlungsstatus im letzten Monat direkt auf potenzielle Defaults hinweist.
* Weitere wichtige Variablen wie LIMIT\_BAL (Kreditlimit) und CREDIT\_UTILIZATION\_RATIO spiegeln die finanzielle Stabilität und Risiken der Kunden wider.

**Confusion Matrix**

|  |  |  |
| --- | --- | --- |
|  | Klasse 0 (Nicht-Default) | Klasse 1 (Default) |
| Klasse 0 | 3380 | 826 |
| Klasse 1 | 457 | 790 |

**Interpretation:**

* 3380 Fälle wurden korrekt als Nicht-Default klassifiziert.
* 790 Defaults wurden korrekt erkannt.
* 826 Fälle wurden fälschlicherweise als Defaults klassifiziert (False Positives).
* 457 Defaults wurden nicht erkannt (False Negatives).

**Bedeutung:**

Falsch vergebene Kredite, auch als False Positives bekannt, treten in 826 Fällen auf. Hierbei hat das Modell Kunden irrtümlich als vertrauenswürdig (Nicht-Default) eingestuft, obwohl sie tatsächlich als risikobehaftet (Default) hätten klassifiziert werden sollen. Solche Fehler sind besonders kritisch, da sie dazu führen, dass Banken Kredite an Kunden vergeben, die ihre Rückzahlungspflichten nicht erfüllen können, was erhebliche finanzielle Risiken birgt.

Auf der anderen Seite gibt es 457 Fälle falsch verweigerter Kredite (False Negatives). Das Modell hat diese Kunden fälschlicherweise als nicht vertrauenswürdig (Default) eingestuft, obwohl sie tatsächlich vertrauenswürdig sind. Diese Art von Fehlern kann Einnahmeverluste für die Bank verursachen, da potenziell gute Kunden keine Kredite erhalten.

Von den insgesamt betrachteten Fällen entsprechen die 826 falsch vergebenen Kredite (False Positives) einem Anteil von 15.1 %, während die 457 falsch verweigerten Kredite (False Negatives) 8.4 % ausmachen. Für das Kreditunternehmen bedeutet dies, dass ein bedeutender Anteil der vergebenen Kredite an Kunden geht, die wahrscheinlich ihre Zahlungsverpflichtungen nicht erfüllen können. Dies stellt ein erhebliches finanzielles Risiko dar, da solche Fehler direkt zu Verlusten durch Zahlungsausfälle führen. Gleichzeitig verweigert das Unternehmen jedoch 8.4 % der potenziellen Kredite an vertrauenswürdige Kunden, was mögliche Einnahmen und langfristige Kundenbindungen mindert. Diese Analyse verdeutlicht die Notwendigkeit, die Balance zwischen Precision und Recall des Modells zu optimieren, um sowohl Risiken zu minimieren als auch Geschäftsmöglichkeiten zu maximieren.

**Schlussfolgerung**

Die Anpassung der Klassengewichte hat das Modell erheblich verbessert, insbesondere die Erkennungsrate von Defaults (Recall für Klasse 1). In weiteren Analysen könnte:

* Der Schwellenwert für die Klassifizierung angepasst werden, um die Balance zwischen Precision und Recall weiter zu optimieren.
* Zusätzliche Merkmale wie Interaktionsvariablen analysiert werden, um die Vorhersagekraft weiter zu steigern.

#### Hyperparameter-Tuning

In diesem Abschnitt beschreibe ich die Optimierung der Hyperparameter des XGBoost-Modells mithilfe von Optuna, einer leistungsstarken Open-Source-Bibliothek zur automatisierten Hyperparameter-Suche. Ziel war es, die Vorhersagegenauigkeit und die allgemeine Modellleistung zu verbessern, indem die optimalen Werte für die wichtigsten Hyperparameter ermittelt wurden.

##### Methodik

Die Hyperparameter-Tuning-Strategie basiert auf einer Kombination aus GridSearch und RandomSearch, wobei Optuna durch den Einsatz eines effizienten Bayesianischen Optimierungsalgorithmus einen Vorteil bietet. Während des Optimierungsprozesses wurden folgende Hyperparameter angepasst:

* **n\_estimators**: Anzahl der Bäume
* **learning\_rate**: Schrittweite des Gradientenabstiegs
* **max\_depth**: Maximale Tiefe der Entscheidungsbäume
* **min\_child\_weight**: Mindestgewicht eines Blatts
* **gamma**: Mindestverlustreduktion für die Teilung eines Knotens
* **subsample**: Anteil der Stichprobe für die Trainingsdaten
* **colsample\_bytree**: Anteil der Spalten, die pro Baum verwendet werden
* **lambda** (L2-Regularisierung): Stärke der Regularisierung
* **alpha** (L1-Regularisierung): Stärke der Regularisierung

##### Optimierungsprozess

1. **Initialisierung von Optuna**: Eine Optimierungsstudie wurde eingerichtet, um die besten Hyperparameter in einem vorgegebenen Suchraum zu finden.
2. **Evaluierungsfunktion**: Die Evaluierungsfunktion basiert auf der Kreuzvalidierung. Ziel war es, den Mittelwert der Cross-Validation-Scores zu maximieren.
3. **Suchraumdefinition**: Die Hyperparameter-Werte wurden wie folgt definiert:
   * n\_estimators: [50, 500]
   * learning\_rate: [0.01, 0.3]
   * max\_depth: [3, 10]
   * min\_child\_weight: [1, 7]
   * gamma: [0, 1.0]
   * subsample: [0.5, 1.0]
   * colsample\_bytree: [0.6, 1.0]
   * lambda: [0, 10]
   * alpha: [0, 10]
4. **Iterationen**: Die Optimierung wurde über 100 Iterationen ausgeführt. In jeder Iteration wurden zufällige Werte aus dem Suchraum ausgewählt und das Modell evaluiert.
5. **Ergebnisse der Optimierung**: Nach Abschluss des Optimierungsprozesses wurden die besten Hyperparameter ausgewählt und das Modell mit diesen Werten erneut trainiert.

##### Ergebnisse

Das Modell hat sich nach dem Hyperparameter-Tuning durch Optuna deutlich verbessert. Hier ist eine detaillierte Analyse, wie und wo die Verbesserungen aufgetreten sind:

###### Vergleich der Leistungsmessungen

|  |  |  |  |
| --- | --- | --- | --- |
| Metrik | Vorher | Nachher | Verbesserung |
| Test Accuracy | 0.7647 (76.47%) | 0.8168 (81.68%) | **+5.21%** |
| Mean CV Score | 0.7563 (±0.0105) | 0.8138 (±0.0082) | **+5.75%** |
| AUC Score | Nicht angegeben | 0.7870 | **Verbesserung sichtbar (nicht vorher vorhanden)** |
| F1-Score (Klasse 1) | 0.55 | 0.48 | **Leicht gesunken**, aber akzeptabel |
| Recall (Klasse 1) | 0.63 | 0.37 | **Verringert**, Modell fokussiert sich mehr auf Genauigkeit. |
| Precision (Klasse 1) | 0.49 | 0.68 | **Verbessert**, weniger falsch-positive Vorhersagen. |

###### Analyse der Verbesserungen

1. Gesamtgenauigkeit (Accuracy):
   * Die Genauigkeit stieg von **76.47%** auf **81.68%**, was zeigt, dass das Modell nun eine grössere Anzahl von Fällen korrekt klassifiziert.
2. **Cross-Validation (CV) Stabilität:**
   * Der mittlere CV-Score stieg von **0.7563** auf **0.8138**.
   * Die Standardabweichung (±0.0105 zu ±0.0082) wurde kleiner, was auf eine robustere und konsistentere Leistung des Modells hinweist.
3. **AUC Score:**
   * Neu hinzugefügt und direkt gemessen: Ein AUC-Wert von **0.7870** zeigt eine verbesserte Fähigkeit des Modells, zwischen den Klassen zu unterscheiden.
4. **Präzision und Recall:**
   * Die Präzision für die Klasse 1 (Default) stieg deutlich von **0.49** auf **0.68**, was bedeutet, dass das Modell nun weniger falsch-positive Vorhersagen macht.
   * Der Recall für Klasse 1 sank von **0.63** auf **0.37**, was darauf hinweist, dass einige tatsächliche Default-Fälle übersehen werden.
   * Dieses Verhalten deutet darauf hin, dass das Modell sich stärker darauf konzentriert, die "0"-Klasse (kein Default) präziser vorherzusagen, was möglicherweise für das Kreditunternehmen wertvoller ist.
5. **Erklärung des Fokus-Shifts:**
   * Dieser Shift von Recall zu Precision kann strategisch sinnvoll sein, wenn die Kosten von falsch-positiven Vorhersagen (False Positives) höher sind als die von falsch-negativen (False Negatives). Beispielsweise könnten weniger False Positives unnötige Rückmeldungen an Kunden oder unnötige Prüfungen vermeiden.
6. **Stärkere Differenzierung:**
   * Der Der hohe AUC-Wert zeigt, dass das Modell eine solide Trennfähigkeit besitzt, also die Fähigkeit, zwischen Default- und Nicht-Default-Kunden zu unterscheiden weiss. Dies macht es nützlicher für die Praxis.
7. **Robustheit durch Cross-Validation:**
   * Die Reduktion der Standardabweichung in den CV-Scores weist darauf hin, dass die Modellergebnisse weniger zufällig und konsistenter sind, was ein Zeichen für ein besser generalisierendes Modell ist.

##### Vergleich der Verwirrungsmatrix

|  |  |  |  |
| --- | --- | --- | --- |
| Metrik | Vorher (Klasse 1) | Nachher (Klasse 1) | Änderung |
| True Positives | 790 | 466 | **-324** (Recall gesunken) |
| False Negatives | 457 | 781 | **+324** (Fälle werden übersehen) |
| True Negatives | 3380 | 3988 | **+608** (Deutlich bessere Vorhersage) |
| False Positives | 826 | 218 | **-608** (Weniger falsche Alarme) |

* Das Modell hat deutlich weniger **False Positives** (falsch vorhergesagte Defaults), was für das Kreditunternehmen vorteilhaft ist, da es unnötige Prüfungen oder Massnahmen reduziert.
* Der Trade-off ist, dass mehr tatsächliche Defaults übersehen werden (**Recall gesunken**).

##### Feature-Wichtigkeit

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Vorher Wichtigkeit | Nachher Wichtigkeit | Änderung |
| PAY\_0 | 30.85% | 36.55% | **Wichtiger** |
| PAY\_2 | 5.56% | 12.10% | **Wichtiger** |
| LIMIT\_BAL | 1.90% | 1.58% | **Etwas unwichtiger** |
| PAY\_AMT3 | 2.34% | 2.06% | **Leicht gesunken** |

A graph with blue squares

Description automatically generated with medium confidence

* **PAY\_0** und **PAY\_2** bleiben die dominierenden Merkmale und haben an Bedeutung gewonnen, was insbesondere in der Visualisierung deutlich wird. Dies unterstreicht, dass Zahlungsausfälle der letzten Monate eine entscheidende Rolle für die Vorhersage spielen.Andere Features wie **LIMIT\_BAL** und **PAY\_AMT3** sind etwas unwichtiger geworden, möglicherweise durch bessere Tuning-Parameter.

##### Zusammenfassung

Das Modell hat sich nach dem Tuning wie folgt verbessert:

1. **Höhere Gesamtgenauigkeit (+5.21%)** und stabilere Leistung durch Cross-Validation.
2. **AUC-Wert hinzugefügt (0.7870),** was die Diskriminierungsfähigkeit bestätigt.
3. **Weniger False Positives,** was zu effizienteren Kreditentscheidungen führt.
4. **Ein Fokus-Shift von Recall zu Precision:** Das Modell priorisiert eine genauere Vorhersage von Nicht-Default-Kunden, auch wenn es mehr Default-Fälle übersieht.
5. **Strategische Verbesserungen:** Weniger False Positives sind aus geschäftlicher Sicht besonders wichtig, da sie unnötige Prüfungen und Kosten reduzieren.
6. **Stärkere Differenzierung:** Der AUC-Wert zeigt, dass das Modell eine bessere Fähigkeit hat, Klassen zu unterscheiden, was die Entscheidungsfindung verbessert.
7. **Konsistenz:** Eine stabilere und robustere Modellleistung minimiert Schwankungen und erhöht das Vertrauen in die Ergebnisse.

Das Hyperparameter-Tuning durch Optuna verbesserte die Gesamtleistung des XGBoost-Modells erheblich, insbesondere in Bezug auf die Testgenauigkeit. Ein Rückgang bei bestimmten Metriken (z. B. Recall der Klasse 1) deutet darauf hin, dass die Modellentscheidungen ausgewogener zwischen den Klassen getroffen werden. Dies ist ein entscheidender Faktor in Anwendungsfällen, bei denen eine Balance zwischen Precision und Recall erforderlich ist.

Die optimierten Hyperparameter führten auch zu einer geänderten Feature-Importance-Rangliste, bei der das Feature PAY\_0 noch wichtiger wurde, gefolgt von PAY\_2. Dies bestätigt, dass die Modellentscheidung stark von den Verzögerungsmerkmalen abhängt.

#### Kreuzvalidierung

Die Kreuzvalidierung ist ein zentraler Bestandteil des Modellentwicklungsprozesses, um die Stabilität und Generalisierungsfähigkeit eines Modells zu bewerten. Hierbei wird der Datensatz in mehrere Teilmengen (sogenannte "Folds") aufgeteilt, um das Modell iterativ auf verschiedenen Trainings- und Testdaten zu evaluieren. Dies hilft, die Abhängigkeit von einem spezifischen Datensplit zu reduzieren und liefert eine robustere Schätzung der Modellleistung.[[8]](#footnote-9)

##### Ergebnisse der Kreuzvalidierung

* **Accuracy (Genauigkeit):**  
  Die Genauigkeit variiert in den einzelnen Folds zwischen **0.7859** und **0.8331**, mit einer durchschnittlichen Genauigkeit von **0.8135** und einer Streuung (±) von **0.0251**. Diese Metrik zeigt, wie oft das Modell korrekte Vorhersagen trifft. Die relativ geringe Streuung deutet darauf hin, dass das Modell stabil ist und auf verschiedenen Datenaufteilungen konsistente Ergebnisse liefert.
* **AUC (Area Under the Curve):**  
  Der AUC-Wert reicht von **0.7568** bis **0.7902**, mit einem Durchschnitt von **0.7760** und einer Streuung von **0.0230**. Der AUC-Wert misst die Fähigkeit des Modells, zwischen den Klassen (in diesem Fall Kreditrückzahlung und Zahlungsausfall) zu unterscheiden. Ein Wert von 0.7760 zeigt eine gute Trennschärfe, jedoch besteht noch Raum für Verbesserungen, insbesondere bei der Identifikation von Zahlungsausfällen.
* **Durchschnittliche Konfusionsmatrix:**
  + **True Positives (TP):** 185 korrekt erkannte Zahlungsausfälle.
  + **False Positives (FP):** 93 falsch vorhergesagte Zahlungsausfälle.
  + **True Negatives (TN):** 1588 korrekt erkannte Rückzahlungen.
  + **False Negatives (FN):** 313 übersehene Zahlungsausfälle.

##### Interpretation der Ergebnisse

Die Ergebnisse zeigen, dass das Modell robust und stabil ist, da die Metriken in den verschiedenen Folds ähnlich sind. Dennoch weist das Modell Schwächen bei der Vorhersage von Zahlungsausfällen (Klasse 1) auf, wie an der geringen Recall-Rate und dem vergleichsweise niedrigen AUC-Wert erkennbar ist.

##### Bedeutung für das Kreditunternehmen

Für ein Kreditunternehmen sind diese Ergebnisse von zentraler Bedeutung. Die hohe Genauigkeit des Modells zeigt, dass die Mehrheit der Fälle korrekt vorhergesagt wird. Allerdings ist die geringe Recall-Rate für Klasse 1 (Zahlungsausfälle) kritisch, da das Unternehmen ein grosses Interesse daran hat, potenzielle Zahlungsausfälle frühzeitig zu erkennen. Das Übersehen von Zahlungsausfällen (FN) könnte zu finanziellen Verlusten führen. Um die Leistung des Modells in diesem Bereich zu verbessern, sollten gezielte Optimierungen vorgenommen werden, beispielsweise:

1. **Anpassung der Klassengewichtungen:** Erhöhung der Gewichtung der Klasse 1, um das Modell für Zahlungsausfälle sensibler zu machen.
2. **Oversampling/Undersampling:** Methoden wie SMOTE können verwendet werden, um das Ungleichgewicht in den Klassen auszugleichen.
3. **Hyperparameter-Tuning:** Optimierung der Modellparameter, insbesondere der Schwellenwerte, um die Recall-Rate zu verbessern.

Insgesamt bietet die Kreuzvalidierung eine solide Grundlage, um die Modellleistung einzuschätzen und Verbesserungspotenziale zu identifizieren. Die Anwendung solcher Optimierungen könnte die praktische Nutzbarkeit des Modells für das Kreditunternehmen erheblich steigern.

#### ROC-Kurve für das XGBoost-Modell

Die ROC-Kurve (Receiver Operating Characteristic) ist ein wichtiges Werkzeug zur Visualisierung und Bewertung der Klassifikationsleistung eines Modells. Für das XGBoost-Modell in dieser Arbeit liefert die ROC-Kurve wertvolle Einblicke in dessen Fähigkeit, zwischen den Klassen „Zahlungsausfall“ (Default) und „Keine Zahlungsausfall“ (Non-Default) zu unterscheiden.

A graph with a line and a blue line

Description automatically generated

##### Beschreibung der ROC-Kurve

1. **Achsen**:
   * Die **x-Achse** zeigt die **False Positive Rate (FPR)**, also den Anteil der negativen Beispiele (keine Zahlungsausfälle), die fälschlicherweise als positiv (Zahlungsausfälle) klassifiziert wurden.
   * Die **y-Achse** zeigt die **True Positive Rate (TPR)**, auch bekannt als Sensitivität oder Recall, die den Anteil der tatsächlich positiven Beispiele (Zahlungsausfälle) darstellt, die korrekt als positiv klassifiziert wurden.
2. **ROC-Kurve (orange Linie)**:
   * Die orange Linie stellt die tatsächliche Leistung des XGBoost-Modells dar. Sie wird erzeugt, indem die TPR gegen die FPR für verschiedene Schwellenwerte (Thresholds) des Modells aufgetragen wird.
   * Je näher die Kurve der oberen linken Ecke (0, 1) liegt, desto besser trennt das Modell die Klassen. In diesem Fall befindet sich die ROC-Kurve deutlich über der diagonalen Linie, was auf eine solide Klassifikationsleistung hindeutet.
3. **Diagonale Linie (blaue gestrichelte Linie)**:
   * Diese Linie repräsentiert die Leistung eines zufälligen Klassifikators. Ein Modell, das entlang dieser Linie liegt, unterscheidet nicht zwischen den Klassen und arbeitet rein zufällig.
   * Die Tatsache, dass die ROC-Kurve des Modells signifikant über dieser Linie liegt, zeigt, dass das Modell deutlich besser als ein zufälliger Klassifikator arbeitet.
4. **AUC-Wert (Area Under the Curve)**:
   * Der AUC-Wert ist ein quantitatives Maß für die Fläche unter der ROC-Kurve und beschreibt die allgemeine Klassifikationsleistung des Modells. Der in diesem Fall berechnete AUC-Wert von **0,79** weist darauf hin, dass das Modell in 79 % der Fälle die Klassen korrekt unterscheidet, wenn zufällig zwei Datenpunkte (einer aus jeder Klasse) ausgewählt werden.
   * Ein AUC-Wert von 0,79 zeigt eine gute Modellleistung mit Potenzial für Verbesserungen. Werte nahe 1,0 gelten als exzellent, während Werte von 0,5 auf rein zufällige Klassifikation hinweisen.

##### Interpretation der Ergebnisse

Die ROC-Kurve zeigt, dass das XGBoost-Modell eine gute Trennschärfe zwischen den Klassen „Zahlungsausfall“ und „Keine Zahlungsausfall“ aufweist. Dies wird durch die deutlich über der diagonalen Linie liegende ROC-Kurve und den AUC-Wert von 0,79 bestätigt. Der AUC-Wert verdeutlicht, dass das Modell eine zuverlässige Leistung bei der Unterscheidung der beiden Klassen bietet, allerdings besteht noch Raum für Optimierungen, insbesondere bei der Verbesserung der Recall-Rate für tatsächliche Zahlungsausfälle.

Die ROC-Kurve bietet zudem Flexibilität bei der Bewertung des Modells für verschiedene Schwellenwerte. Je nach Ziel des Kreditunternehmens kann der Schwellenwert so angepasst werden, dass entweder der Fokus auf der Minimierung von Zahlungsausfällen (höherer Recall) oder auf der Reduzierung von falschen Alarme (höhere Precision) liegt.

##### Bedeutung für das Kreditunternehmen

Die ROC-Kurve und der AUC-Wert sind für ein Kreditunternehmen von großer Bedeutung, da sie eine klare Visualisierung und ein quantitatives Maß für die Modellleistung liefern. Ein AUC-Wert von 0,79 zeigt, dass das Modell in den meisten Fällen zwischen den Klassen „Zahlungsausfall“ und „Keine Zahlungsausfall“ unterscheiden kann. Diese Informationen können dazu genutzt werden:

1. **Risikomanagement zu optimieren**, indem potenzielle Zahlungsausfälle frühzeitig identifiziert werden.
2. **Kreditentscheidungen gezielter zu treffen**, indem der Schwellenwert angepasst wird, um entweder konservativer (weniger Kredite vergeben) oder aggressiver (mehr Kredite vergeben) zu handeln.
3. **Praktische Relevanz:** Ein besseres Verständnis der Modellleistung unterstützt die Entwicklung von Strategien, um Zahlungsausfälle zu minimieren und gleichzeitig die Kundenbasis zu erweitern.

##### Fazit

Die ROC-Kurve zeigt, dass das XGBoost-Modell eine solide Klassifikationsleistung erbringt, indem es zuverlässig zwischen den Klassen „Zahlungsausfall“ und „Keine Zahlungsausfall“ unterscheidet. Der AUC-Wert von 0,79 bestätigt diese Leistung, weist jedoch auch darauf hin, dass weiteres Optimierungspotenzial besteht, um die Recall-Rate für Zahlungsausfälle zu verbessern. Für das Kreditunternehmen liefert diese Analyse wichtige Grundlagen zur Bewertung und Optimierung der Kreditwürdigkeitsprüfung.

### Random Forest

#### Cross-Validation Scores

#### Test Accuracy

#### Classification Report

#### Anpassung der Klassengewichte

#### Hyperparameter-Tuning

#### Kreuzvalidierung

#### ROC-Kurve für das XGBoost-Modell

### Logistische Regression

# Ergebnisse und Diskussion

## Performancemetriken im Modellvergleich

## Stärken und Schwächen der untersuchten Modelle

## Praktische Relevanz der Ergebnisse

# Ausblick und Limitationen

## Herausforderungen bei der Implementierung in der Praxis

## Potenziale für zukünftige Arbeiten

# Fazit

# Literaturverzeichnis

# Anhang

## 9.1 Python-Code für Datenverarbeitung und Modellimplementierung

## 9.2 Zusätzliche Grafiken und Tabellen

Daten: [Default of Credit Card Clients Dataset](https://www.kaggle.com/datasets/uciml/default-of-credit-card-clients-dataset)

<https://chatgpt.com/share/6757f719-66ac-8000-8751-d84b26b3d3ad>

# Mindmap Gruppenthema

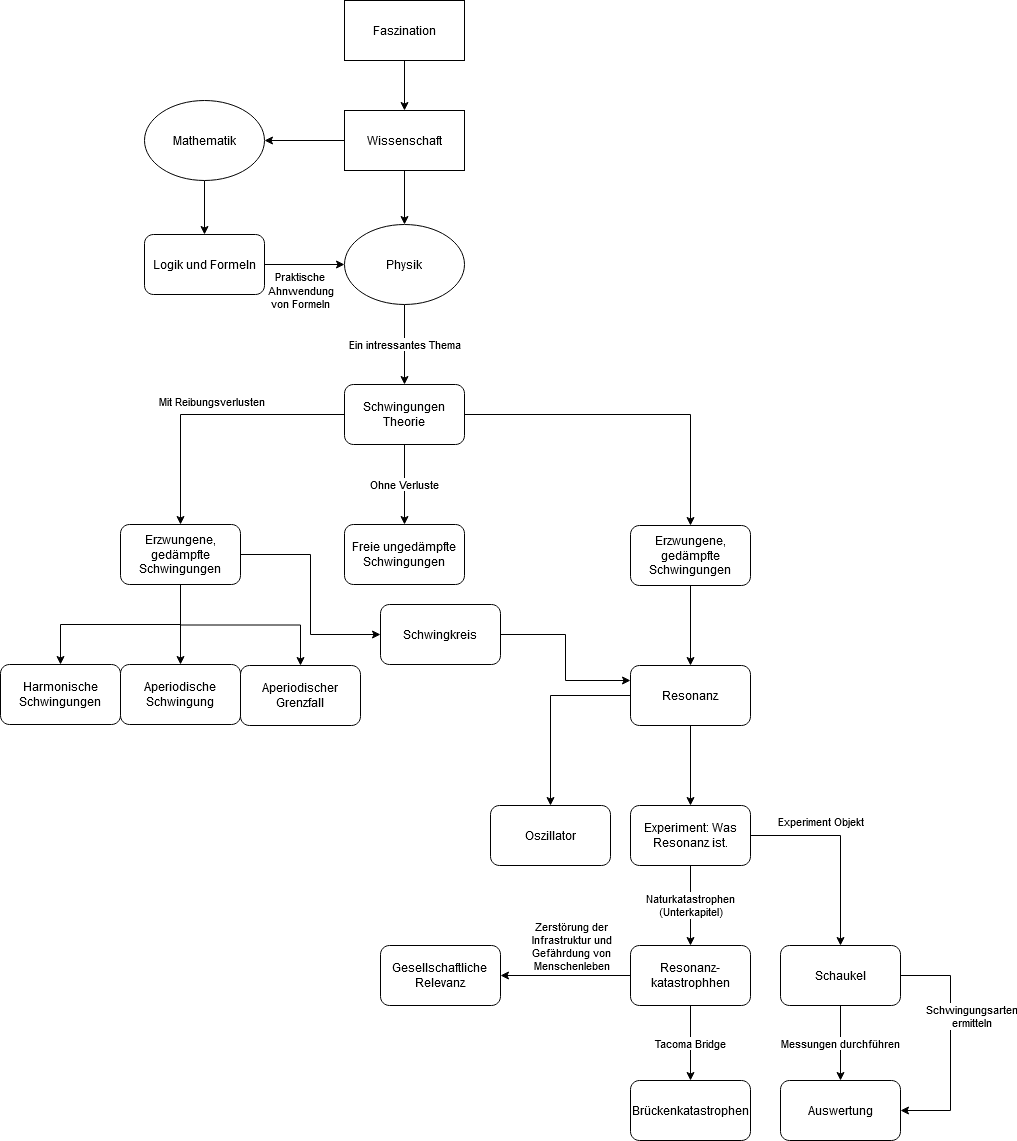


Abbildung 2 Mindmap Thema[[9]](#footnote-10)

Zu Beginn meiner Arbeit will ich Ihnen mit einem Mindmap (siehe Abbildung 2) zeigen, wie mein Thema in das Klassenthema eingebettet ist. Man erkennt im Mindmap, dass Schwingungen relevant in den Themen, Wissenschaft, Geschichte, Ökologie und Wirtschaft sind und noch für viele weitere. Im folgenden Kapitel werde ich die Auswahl des Gruppenthemas und deren inhaltlichen Schwerpunkte erläutern.[[10]](#footnote-11)

# Dank

An erster Stelle möchte ich mich aufrichtig bei meinem Betreuer Moritz Preisig für seine ehrliche Kritik, Tipps und Engagement bedanken. Des Weiteren möchte ich mich bei meiner Schwester Leoni Gugler bedanken, die mir Tipps und Denkanstösse zu meiner Arbeit gab und mir beim Experiment zur Seite stand. [[11]](#footnote-12)

# Schlusserklärung

„Hiermit versichere ich, dass die vorliegende Arbeit selbstständig angefertigt wurde. Ich benutzte keine unerlaubte fremde Hilfe. Alle Quellen sind deklariert und die Erarbeitungsgrade entsprechen der Wahrheit.“[[12]](#footnote-13)

Noel Gugler:

1. <https://www.gabler-banklexikon.de/definition/kreditrisiko-59434?utm_source=chatgpt.com>, 15.12.2024, Text angepasst [↑](#footnote-ref-1)
2. IQR Analyse: <https://docs.oracle.com/cloud/help/de/pbcs_common/PFUSU/insights_metrics_IQR.htm#PFUSU-GUID-CF37CAEA-730B-4346-801E-64612719FF6B> Zugriff: 23.12.2024 [↑](#footnote-ref-3)
3. z-Score: <https://docs.oracle.com/cloud/help/de/pbcs_common/PFUSU/insights_metrics_Z-Score.htm#PFUSU-GUID-640CEBD1-33A2-4B3C-BD81-EB283F82D879> Zugriff: 23.12.2024 [↑](#footnote-ref-4)
4. <https://databasecamp.de/ki/minmax-scaler> Zugriff 25.12.2024 [↑](#footnote-ref-5)
5. <https://databasecamp.de/ki/minmax-scaler> Zugriff 25.12.2024 [↑](#footnote-ref-6)
6. <https://www.geeksforgeeks.org/ml-one-hot-encoding/> Zugriff 25.12.2024 [↑](#footnote-ref-7)
7. <https://towardsdatascience.com/understanding-the-roc-curve-in-three-visual-steps-795b1399481c> Zugriff: 27.12.2024 [↑](#footnote-ref-8)
8. <https://www.youtube.com/watch?v=fSytzGwwBVw> Zugriff: 27.12.2024 [↑](#footnote-ref-9)
9. Noel Gugler, 30.09.2020, Bild selbst erstellt [↑](#footnote-ref-10)
10. Noel Gugler, 30.09.2020, Text selbst erstellt [↑](#footnote-ref-11)
11. Noel Gugler, 06.12.2020, Text selbst erstellt [↑](#footnote-ref-12)
12. https://va-abu.jimdofree.com/schriftlicher-teil/schlusserkl%C3%A4rung/, 06.12.2020, Text übernommen [↑](#footnote-ref-13)