



University of Applied Science - Fernstudium

Studiengang: Master Data Science (FS MADW-120 2022 WS)

Portfolio Phase 3

Projekt: Data Science Use Case (DLMDWPDSUC01)

Autor: Benno Verschueren
Matrikelnummer: IU14076010

Tutor: Dr. Thomas Zöllner

Datum: 16.06.2024

I Abstract

In dieser Arbeit wird die Implementierung eines automatisierten Echtzeit-Meldesystems (AEMS) für das fiktive Game Studio RagePlayer® vorgestellt, das zur Prävention von Cybermobbing in In-Game-Chats entwickelt wurde. Das System basiert auf einer Sentimentalanalyse und verwendet ein KI-basiertes Klassifikationsmodell. Nachrichten, die gegen die Richtlinien von RagePlayer® verstoßen, werden automatisch gekennzeichnet und unleserlich dargestellt. Ein zentrales Ziel des AEMS ist es, eine sichere und positive Spielumgebung zu fördern, indem schädliches Verhalten in Echtzeit erkannt und entsprechend gehandhabt wird. Neben der Kennzeichnung und Unlesbarmachung von beleidigenden Nachrichten informiert das System Moderatoren automatisch über konkrete Fälle von Cybermobbing, was die Effizienz des Moderationsteams erhöht und Kosten reduziert. Zusätzlich werden Daten zu Verstößen protokolliert, um langfristige Analysen und Verbesserungen des Systems zu ermöglichen sowie potenzielle rechtliche Schritte gegen Verstöße zu unterstützen.

Die Planbarkeit und Umsetzung von dem AEMS wird durch den Einsatz des Machine Learning Canvas (MLC) strukturiert und optimiert. Der MLC dient als Framework, das die verschiedenen Aspekte der Entwicklung und Implementierung des AEMS, wie zum Beispiel die Datenbeschaffung, Vorverarbeitung, Modelltraining und Evaluierung, systematisch abbildet. Dies ermöglicht eine effektive und transparente Entwicklung, die sowohl technische als auch organisatorische Herausforderungen adressiert. Durch den Einsatz des MLC konnten bereits in der Planungsphase potenzielle Aufwände für die Beschaffung gekennzeichnetener Trainings- und Testdaten identifiziert werden. Zudem ermöglichte der ML-Canvas eine grobe Abbildung der Neustrukturierung des Geschäftsprozesses von RagePlayer® durch die Integration des AEMS. Abschließend werden Evaluierungsmaßnahmen anhand wichtiger Key Performance Indicators (KPIs) durchgeführt, um den potenziellen Einsatz des AEMS umfassend zu bewerten und zu optimieren.

Inhaltsverzeichnis

I	Abstract	II
II	Abbildungsverzeichnis	V
III	Abkürzungsverzeichnis	VI
1	Value Proposition	1
1.1	Einleitung	1
1.2	Why?	1
1.3	What?	2
2	Machine Learning Task	3
2.1	Sentimentalanalyse	3
2.2	Maschinelles Lernen	3
3	Decisions	4
3.1	Integration von dem AEMS bei RagePlayer®	4
4	Making Predictions	5
4.1	Latenz in Chatsystemen	5
5	Offline Evaluation	6
5.1	Evaluierung von Klassifikationsmodelle	6
5.2	Bewertungskriterien	6
6	Data Sources	7
6.1	Interne Datenquelle	7
6.2	Trainings- und Testdaten	7
7	Collection Data	8
7.1	Kontextuelle Daten	8
7.2	Textdaten	8
8	Features	9
8.1	Textinhalt (Feature Extraction)	9
9	Building Models	10
9.1	Auslöser für erneutes Training des ML-Modells	10
10	Live Evaluation and Monitoring	11
10.1	Churnrate	11
10.2	Customer Satisfaction Score (CSAT)	11

10.3 A/B Test	11
11 ML-Canvas	12
12 Pitchdeck	13

II Abbildungsverzeichnis

1	Schematische Darstellung von dem AEMS	2
2	Grafische Prozessdarstellung von dem AEMS	4
3	Aufteilung der Latenzen	5
4	Beispiel für Chat mit Melde-Button	7
5	ML-Canvas	12
6	Pitchdeck Folie 1	13
7	Pitchdeck Folie 2	13
8	Pitchdeck Folie 3	14
9	Pitchdeck Folie 4	14
10	Pitchdeck Folie 5	15

III Abkürzungsverzeichnis

AEMS	Automatisches Echtzeit Meldesystem
ASCII	American Standard Code for Information Interchange
CSF	Critical Success Factor
CSAT	Customer Satisfaction Score
DotA	Defense Of The Ancients - eine Multiplayer Online Battle Arena bzw. ein Action-Echtzeit-Strategiespiel
KI	Künstliche Intelligenz
KPI	Key Performance Indicator
ML	Machine Learning
MLC	Machine Learning Canvas
MMORPG	Massively Multiplayer Online Role-Playing Game
NLP	Natural Language Processing
SQL	Structured Query Language
SVM	Support Vector Machine

1 Value Proposition

1.1 Einleitung

Das fiktive Game Studio RagePlayer® hat mit der Veröffentlichung ihres ersten Spiels „Space Gladiators 4096“, ein futuristischer Third-Person-Shooter, einen beachtlichen Erfolg in der Gaming-Szene erzielt. Durch das Online-Spielkonzept können Gamer virtuell in Teams gegeneinander antreten. Um das kooperative Gameplay zu fördern, ermöglicht der bereitgestellte In-Game-Chat den Gamern, einfach und schnell mit anderen Spielern in Kontakt zu treten. Mit der steigenden Nutzerzahl kommt es jedoch in diesen Chats zunehmend zu Fällen von Cybermobbing, die bereits negative mediale Aufmerksamkeit erregt haben.

1.2 Why?

Problem

Cybermobbing ist ein Verhalten, bei dem absichtlich und wiederholt über elektronische Geräte psychischer Schaden an anderen Personen verursacht wird. Zu den Folgen von Cybermobbing gehören unter anderem soziale Isolation, psychische Störungen und im schlimmsten Fall Suizid (Murnion et al. 2018, S. 1). Neben den klassischen digitalen Kanälen wie Chats, Foren, E-Mails und diversen Social-Media-Plattformen sind auch sogenannte In-Game-Chats von der Problematik betroffen. Zu den häufigsten Formen der Online-Belästigung unter Gamern gehören Trolling (64%), Hassreden (57%) und persönliche Drohungen (47%) (Petrov (2024)).

Eine Studie zeigt, dass 28% der Online-Multiplayer-Gamer, die im Spiel Belästigungen erfahren haben, bestimmte Spiele aufgrund ihres Rufes für feindselige Umgebungen meiden. Zudem gaben 22% an, dass sie ganz aufgehört haben, bestimmte Spiele zu spielen. Besonders betroffen sind jene Spiele, die als besonders toxisch wahrgenommen werden und in denen Spieler verstärkt auf feindselige Interaktionen treffen. Diese Spiele umfassen unter anderem „DotA 2“ (40%), „Call of Duty“ (39%), „League of Legends“ (36%), „Fortnite“ (35%), „Grand Theft Auto“ (34%) und „Valorant“ (33%). In diesen Fällen gaben die meisten Online-Multiplayer-Spieler an, entweder vorsichtiger geworden zu sein oder ganz aufgehört zu haben, sie zu spielen. (ADL (2020)). Das Spiel „Space Gladiators 4096“ von RagePlayer® kann demselben Genre zugeordnet werden wie die oben genannten Spiele.

Negative Auswirkungen für das Game Studio

Neben den bereits genannten psychischen Belastungen für die Gamer führt Cybermobbing auch zu erheblichen ökonomischen Problemen für RagePlayer® (getstream.io (2022)):

- Schädigung des Markenrufs
- Hohe Abwanderungsraten bei Spielern
- Verlust von Partnerschaften
- Beendigung des Werbesponsorings im Spiel

Ziel

Proaktive Maßnahmen zur Vermeidung von Cybermobbing erhöhen das Vertrauen und die Kundenbindung der Gamer sowie gegebenenfalls auch deren Eltern. Ein weiterer positiver Aspekt ist die medienwirksame und klare Positionierung von RagePlayer® gegen Hass im Netz. Einige Computerspieleanbieter setzen bereits Verhaltens- und Kommunikationsrichtlinien ein, um Cybermobbing zu bekämpfen. Diese Maßnahmen sind jedoch nur ein Teil der Gesamtlösung.

1.3 What?

Für RagePlayer® soll ein automatisiertes Echtzeit-Meldesystem (AEMS) implementiert werden, um Cybermobbing in den In-Game-Chats zu verhindern. Basierend auf der Sentimentanalyse wird ein KI-basiertes Klassifikationsmodell Nachrichten automatisch kennzeichnen und unleserlich darstellen, wenn sie gegen die Richtlinien von RagePlayer® verstoßen. Das AEMS leitet die relevanten Informationen gleichzeitig an die Moderatoren weiter. Zusätzlich kann durch die Automatisierung wiederkehrender und zeitaufwändiger Kontrollaufgaben die Arbeitskosten erheblich gesenkt werden. Darüber hinaus ermöglicht ein automatisiertes System die Skalierbarkeit, um mit der wachsenden Nutzerzahl Schritt zu halten.

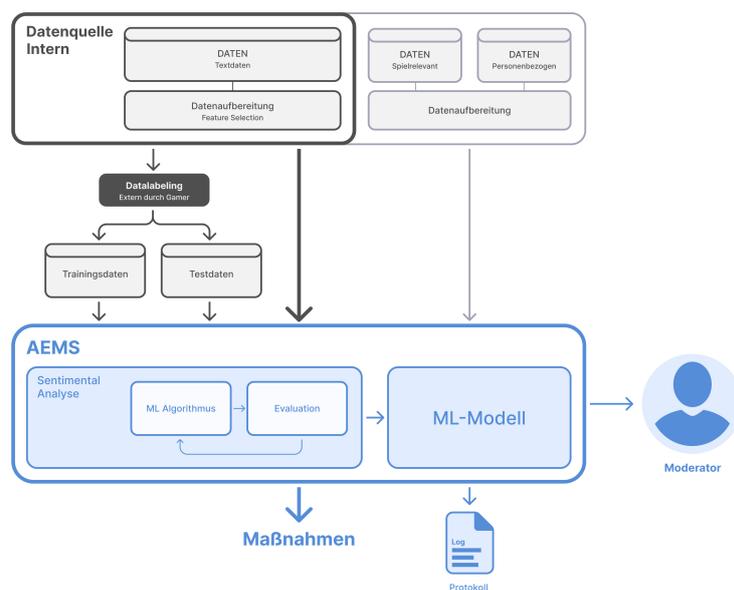


Abbildung 1: Schematische Darstellung von dem AEMS

2 Machine Learning Task

Die Textinhalte sind in toxische und nicht-toxische Kategorien zu klassifizieren, wobei überwachte Lernalgorithmen zum Einsatz kommen. Darüber hinaus ist eine Sentimentanalyse erforderlich, um die emotionale Tonalität der Texte zu bewerten.

2.1 Sentimentanalyse

Die Sentimentanalyse, ist ein Verfahren zur Untersuchung von Wörtern in ihrem semantischen Kontext. Da Wörter oft mehrere Bedeutungen haben können, ist es erforderlich, ihre tatsächliche Bedeutung innerhalb eines Satzes zu erfassen, indem die umgebenden Wörter als Ganzes betrachtet werden (datascientest.com (2023)).

Arten der Sentimentanalyse

Die Methoden der Sentimentanalyse lassen sich generell in drei Hauptansätze gliedern:

- Maschinelles Lernansatz
- Lexikonbasierter Ansatz
- Hybrider Ansatz

Beim maschinellen Lernansatz (ML) werden gängige ML-Algorithmen angewendet, die auf linguistischen Merkmalen basieren. Der lexikonbasierte Ansatz hingegen greift auf ein Sentiment-Lexikon zurück, das eine Zusammenstellung bekannter und vordefinierter Sentiment-Begriffe darstellt. Dieser Ansatz lässt sich weiter in einen wörterbuchbasierten und einen korpusbasierten Ansatz unterteilen, wobei statistische oder semantische Methoden genutzt werden, um die Stimmungspolarität zu bestimmen. Der hybride Ansatz vereint beide Ansätze und ist weit verbreitet, wobei Sentiment-Lexika eine wesentliche Rolle in den meisten Methoden spielen (Wnkhade et al. 2022, S.5744).

2.2 Maschinelles Lernen

Support Vector Machines (SVM)

Aufgrund der hohen Leistungsfähigkeit und Genauigkeit wird der überwachte Lernalgorithmus Support Vector Machines (SVM) für den Einsatz von Sentimentklassifikation bevorzugt (Singh / Jaiswal (2023)). Ein weiterer großer Vorteil von SVM ist, dass dieses Modell auch mit begrenzten Datenmengen präzise Ergebnisse liefert. Um die Annahme zu stützen, kann der Klassifikator durch mehrere Bewertungsmetriken evaluiert werden. Experimente bestätigen, dass die SVM gegenüber anderen Ansätzen, z. B. Naive Bayes, in der Textanalyse genauer arbeiten (Atoum 2020, S.295).

3 Decisions

3.1 Integration von dem AEMS bei RagePlayer®

Durch die Automatisierung des Ablaufs von dem AEMS können die Personalkosten reduziert werden. Moderatoren werden erst dann durch das System benachrichtigt, wenn ein konkreter Fall von Cyberbullying vorliegt. Nach der Identifizierung eines Verstoßes in einem In-Game-Chat durch das AEMS werden die folgenden Schritte im Geschäftsprozess automatisch eingeleitet:

- **Schritt 1: Kennzeichnung und Filterung von toxischen Nachrichten**
Im In-Game-Chat werden Verstöße durch eine unleserliche Darstellung gekennzeichnet, um sowohl den Verfasser als auch den Empfänger der Nachrichten visuell über den Verstoß zu informieren.
- **Schritt 2: Warnung und Sperrung**
Im Falle eines Verstoßes wird dem Verfasser eine Warnung übermittelt. Bei wiederholten Verstößen wird die Chatfunktion für den Verfasser automatisch deaktiviert.
- **Schritt 3: Automatisierte Benachrichtigung** Das AEMS generiert automatisch Benachrichtigungen an Moderatoren, die detaillierte Informationen über den erkannten Verstoß sowie ergänzende Daten zum Gesprächsverlauf enthalten.
- **Schritt 4: Protokollierung**
Aufgrund der Tatsache, dass Cybermobbing als Straftatbestände gemäß dem Strafgesetzbuch (StGB) relevant werden können (www.bundestag.de (2016)), werden Vorfälle protokolliert und gespeichert.

Grafische Darstellung des integrierten Prozessablaufs

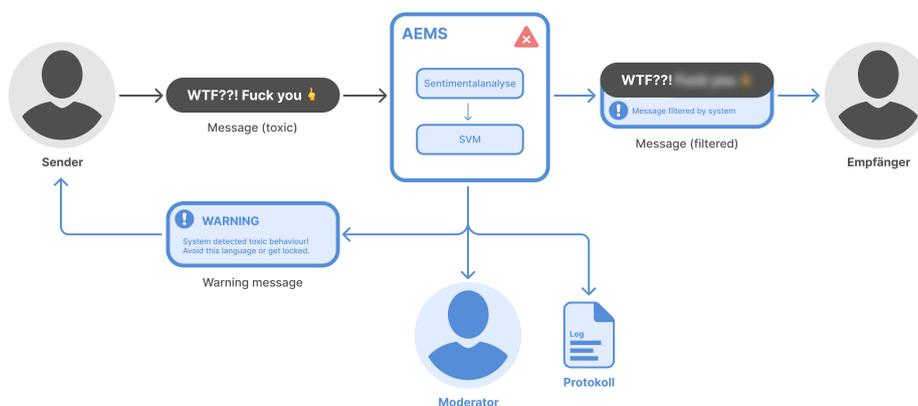


Abbildung 2: Grafische Prozessdarstellung von dem AEMS

4 Making Predictions

Das ML-Modell für die Sentimentanalyse erfordert die Implementierung eines Echtzeitsystems. Die Echtzeitberechnung für die Features stellt eine entscheidende Komponente für die Anwendung dar. In diesen Szenarien ist eine minimale Latenzzeit (gemessen in Millisekunden) bei der Berechnung neuer Datenmerkmale von entscheidender Bedeutung, um präzise und hochwertige Schlussfolgerungen zu ermöglichen (Zhang et al. 2023, S. 77).

4.1 Latenz in Chatsystemen

Latenz bezeichnet die zeitliche Verzögerung zwischen Eingabe und Ausgabe und ist ein wesentlicher Aspekt der Kommunikation zwischen Mensch und Computer. Für Chatsysteme wird eine Reaktionszeit zwischen 200ms und 1000ms empfohlen, um den Erwartungen der Benutzer gerecht zu werden (Chen et al. 2004, S.64). Ausgehend von diesen Richtwerten wird eine maximale Bearbeitungszeit von 600ms für das AEMS festgelegt. Daraus resultieren folgende Zeitaufteilungen für den gesamten Prozess:

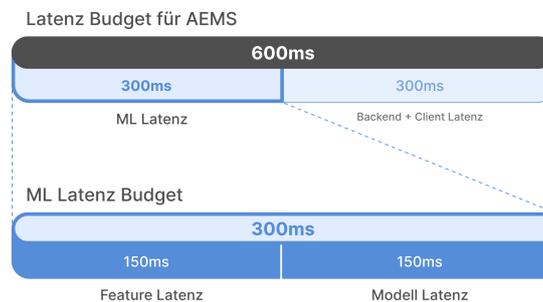


Abbildung 3: Aufteilung der Latenzen für das AEMS (in Anlehnung an Simba (2024))

5 Offline Evaluation

5.1 Evaluierung von Klassifikationsmodelle

Überwachtes maschinelles Lernen bietet verschiedene Bewertungsmethoden zur Leistungsbeurteilung eines Modells. Da die Genauigkeit anwendungsabhängig ist, gibt es keinen allgemeinen Schwellwert für gute Performance. Genauigkeitswerte zwischen 70% und 90% gelten jedoch oft als ideal und entsprechen vielen Industriestandards (Barkved (2022)).

5.2 Bewertungskriterien

Zur Bewertung eines Klassifikationsmodells ist die Einführung eines Qualitätsmaßes erforderlich (Arkhipov 2018, S. 22). Bei der Bewertung von NLP Modellen werden häufig folgende Metriken bevorzugt (Sokolova et al. (2006)):

Genauigkeit

Die Genauigkeit, auch Spezifität genannt, ist wichtig, um sicherzustellen, dass die meisten der als Cybermobbing markierten Fälle tatsächlich Cybermobbing sind. Dies reduziert die Anzahl der Fehlalarme (False Positives), die unnötige Maßnahmen zur Folge haben könnten.

$$Precision = \frac{TP}{TP + FP} \quad (5.1)$$

Trefferquote

Die Trefferquote, auch Sensitivität genannt, ist wichtig, um sicherzustellen, dass möglichst wenige Fälle von Cybermobbing übersehen werden. Dies ist entscheidend für den Schutz der Nutzer.

$$Recall = \frac{TP}{TP + FN} \quad (5.2)$$

F1-Score

Für die Bewertung von dem AEMS wird der F1-Score verwendet, da er sowohl die Präzision als auch die Trefferquote in einem einzigen gewichteten Mittelwert zusammenfasst. Dies ist besonders wichtig, da sowohl die Vermeidung von Fehlalarmen (hohe Präzision) als auch die Erfassung aller echten Fälle (hoher Recall) entscheidend sind (Wnkhade et al. 2022, S.5762).

$$F1-Score = \frac{2(Precision \cdot Recall)}{Precision + Recall} \quad (5.3)$$

6 Data Sources

6.1 Interne Datenquelle

Die Daten für das AEMS werden vollständig durch interne Quellen abgedeckt. Neben den Textdaten, die zur Auswertung des ML-Modells herangezogen werden, werden auch personenbezogene und spielrelevante Daten zu Protokollierungszwecken (siehe 3.1) erfasst.

6.2 Trainings- und Testdaten

Für das Training und die Evaluierung des ML-Modells müssen die internen Daten für die Trainings- und Testphasen entsprechend aufbereitet und gekennzeichnet werden. Es ist entscheidend, Kriterien festzulegen, die einen Verstoß eindeutig definieren. Das sogenannte Labeling ist ein zeit- und ressourcenintensiver Prozess, der typischerweise manuell durchgeführt wird. Um die mit dem Labeling verbundenen Kosten zu reduzieren, wird im bestehenden In-Game-Chat eine neue Funktion eingeführt. Diese Funktion ermöglicht es den Spielern, Chatnachrichten anderer Spieler als anstößig zu markieren, indem sie einen Melde-Button betätigen. Diese Nutzerberichte können anschließend als zusätzliche Datenquelle für das Labeling verwendet werden, wodurch der Aufwand für die manuelle Kennzeichnung verringert und die Effizienz des Prozesses gesteigert wird.

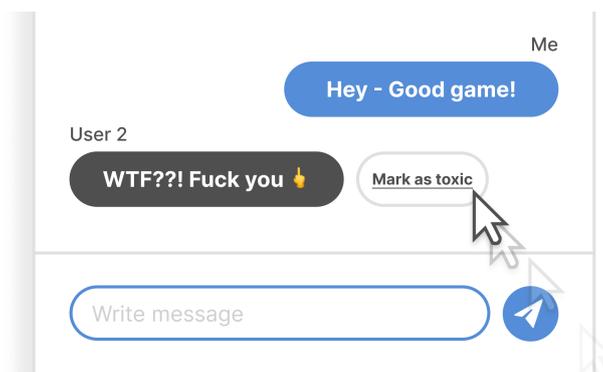


Abbildung 4: Beispiel für Chat mit Melde-Button

7 Collection Data

7.1 Kontextuelle Daten

Zu Werbezwecken, Marktforschung und der bedarfsgerechten Spielgestaltung (z.B. Verbesserung von Teambildung im Spiel, durch Lokalisierung von Spielern) werden bei der Registrierung auf der RagePlayer® Plattform personenbezogene Daten erhoben. Diese Daten werden gemäß den geltenden Datenschutzbestimmungen (ULD 2010, S. 42) in einer SQL-Datenbank gespeichert und verarbeitet. Die erhobenen Informationen werden für die Protokollierung von möglichen Verstößen verwendet (siehe 3.1).

7.2 Textdaten

Der Textinhalt bildet den Hauptbestandteil der Analyse und wird aus einem unstrukturierten Datensatz als interne Datenquelle bezogen. Bevor die Features definiert werden können, ist eine Aufbereitung des Textinhalts erforderlich. Dabei werden verschiedene Vorverarbeitungsschritte, auch bekannt als Data Preprocessing, durchgeführt (Shahana / Omman 2015, S.1587):

- Umwandlung Groß- in Kleinbuchstaben
- Normalisierung (z. B. Entfernung von Apostrophe)
- Entfernung von Nicht-ASCII Zeichen, Leerzeilen, Satzzeichen und Stoppwörter (z.B. Artikel)
- Stemming (Zusammenfassung von verwandten Wörter)

8 Features

Für das AEMS müssen spezifische Featurevariablen bestimmt werden. Diese Variablen dienen als Eingangsparameter für das Modell und beeinflussen maßgeblich die Leistungsfähigkeit.

8.1 Textinhalt (Feature Extraction)

Forschungsarbeiten haben das Kommunikationsverhalten in Ego-Shooter-Videospielen untersucht ((Herring et al. 2009, S.5)). Die Chatnachrichten in solchen Spielen sind in der Regel kurz und enthalten häufig Emoticons sowie einen spezifischen Jargon des Spiels. Der Gebrauch von Emoticons in Spielen kann verschiedene funktionale Rollen erfüllen, wie beispielsweise im Fall von „Fuck you :)“, wo das Emoticon die Botschaft abschwächt und darauf hinweist, dass sie scherzhaft gemeint ist (Thompson et al. 2017, S.7). Auch bei „Space Gladiators 4096“ findet dieses Kommunikationsverhalten, einschließlich des erwähnten spielspezifischen Jargons, Anwendung und muss dementsprechend berücksichtigt werden. Zur nachfolgenden Stimmungsklassifizierung müssen aus den Chatnachrichten Textmerkmale extrahiert werden. Dieser Prozess ist besser bekannt als „Feature Extraction“ (Medhat et al. 2014, S. 1095):

Frequenz

Die Häufigkeit bestimmter Wörter im Text kann Aufschluss darüber geben, welche Wörter mit Cybermobbing assoziiert werden.

N-Gramme

Die Verwendung von N-Grammen (Sequenzen von aufeinanderfolgenden Wörtern) kann helfen, Kontext und Bedeutung im Text besser zu verstehen. Es gibt den Wörtern entweder eine binäre Gewichtung (Null, wenn das Wort vorkommt, oder Eins, wenn nicht) oder verwendet Begriffshäufigkeitsgewichtungen, um die relative Bedeutung von Merkmalen anzugeben (Swati / Rajesh 2015, S.635).

Emotionale Wörter

Für die Analyse müssen emotionale Wörter und Ausdrücke definiert werden. Diese liefern wichtige Hinweise, auf die vorherrschende Stimmung im Text. Neben dem reinen Textinhalt müssen auch Emojis und Emoticons berücksichtigt werden, da sie häufig verwendet werden, um Emotionen zu vermitteln und die Bedeutung einer Nachricht verstärken.

9 Building Models

Durch regelmäßiges erneutes Trainieren (Retraining) des ML-Modells mit neuen Daten können potenzielle Daten- und Konzeptdrifts vermieden werden. Somit kann die Leistungsfähigkeit des AEMS konstant gehalten bzw. verbessert werden.

Datendrift

Dieser Umstand tritt vor allem auf, wenn sich die Daten und deren Eigenschaften im Laufe der Zeit ändern und sich somit die Produktionsdaten von den Trainingsdaten stark unterscheiden (Dilmegani (2024)). Bei „Space Gladiators 4096“ könnte dieser Effekt auftreten, wenn das Spiel bzw. die Funktionalitäten durch Updates sich ändern und dadurch auch das Kommunikationsverhalten der Gamer.

Konzeptdrift

entsteht, wenn sich die Beziehung zwischen Eingangs- und Ausgangsvariablen ändert und somit prädiktive Leistung des ML Algorithmus verringert (Dilmegani (2024)). Da sich das Kommunikationsverhalten bei den Gamern langsam bzw. kaum ändern, muss auf diesen Umstand nicht wesentlich eingegangen werden. Ein erneutes Trainieren ist ein kostspieliger Prozess, da das ML-Modell in dieser Phase nicht einsatzfähig ist. Zusätzlich müssen nach dem Training Evaluierungsmaßnahmen durchgeführt werden, um die Leistungsfähigkeit zu überprüfen. Der Zeitpunkt eines neuerlichen Trainierens kann unter anderem durch folgende Auslöser durchgeführt werden (Sanjay (2022)).

9.1 Auslöser für erneutes Training des ML-Modells

Da sich das Kommunikationsverhalten bzw. die Daten für die semantischen Analyse nicht dynamisch ändern, werden folgende Faktoren den Zeitpunkt eines Neustrainings bestimmen:

Bedingung 1: Updates bei „Space Gladiators 4096“

Die Einführung neuer Missionen, Funktionen oder Spielmechaniken kann zu neuen Kommunikationsmustern und Vokabularen führen, die im ML-Modell berücksichtigt werden müssen. Mit der Zeit entwickeln sich neue Trends und Slangs, die in den Chats verwendet werden. Das Modell muss diese erkennen und entsprechend reagieren können.

Bedingung 2: Unterschreitung von Leistungsschwellwerten

Wenn während der Evaluierung Leistungsschwellwerte unterschritten werden, die anhand manuell markierter Verstöße und automatischer Markierungen festgelegt wurden.

10 Live Evaluation and Monitoring

Für eine erfolgreiche Marktablierung von RagePlayer® muss das Studio Alleinstellungsmerkmale anbieten, die als "kritische Erfolgsfaktoren"(CSF) bezeichnet werden. CSFs definieren die Performanz und Leistungsfähigkeit von RagePlayer® (Parmenter, S.161). Man unterscheidet zwischen operativen Erfolgsfaktoren, die interne Prozesse betreffen, und externen Faktoren, die die Wahrnehmung der Kunden in den Vordergrund stellen (Parmenter, S.162). Das AEMS soll das Vertrauen und die Bindung der Spieler sowie ihrer Eltern stärken und ist daher als externer Erfolgsfaktor zu betrachten. Um die Kundenbindung messbar zu machen, müssen „Key Performance Indicators“ (KPI) definiert werden, die messbare Metriken ermöglichen. Die so genannte Abwanderungsrate (Churnrate) ist eine Möglichkeit die Kundenbindung zu interpretieren (Bernazzani Barron (2023)).

10.1 Churnrate

Die Kundenabwanderungsrate ist der Prozentsatz der Kunden, bzw. Gamer, die das Spiel dauerhaft verlassen. Diese Entscheidung kann bewusst oder unbewusst sein und aus externen oder internen Gründen erfolgen (Runge et al. (2014)). Eine niedrige Churnrate bedeutet eine starke Kundenbindung.

$$\text{Churnrate} = \left(\frac{\text{Lost Customers}}{\text{Total Customer at a Start of Time Period}} \right) \cdot 100 \quad (10.1)$$

10.2 Customer Satisfaction Score (CSAT)

Der CSAT kann durch Umfragen ermittelt werden, bei denen die Gamer gebeten werden, ihre Zufriedenheit mittels einer Bewertungsskala zu bewerten. Es ist jedoch wichtig sicherzustellen, dass diese Umfragen nicht als störend für den Spieler empfunden werden. Für die Berechnung des CSAT muss die Bewertungsskala in positive und negative Bewertungen unterteilt werden (Leiba et al. 2023, S.108).

$$\text{CSAT} = \left(\frac{\text{Number of positive responses}}{\text{Total Number of responses}} \right) \cdot 100 \quad (10.2)$$

10.3 A/B Test

A/B-Test ist eine Methode, bei der zwei Varianten, verglichen werden, um die „bessere“ Variante zu ermitteln (Siroker / Koomen (2013)). Es wird der berechnete KPI vor und nach der Implementierung der AEMS ermittelt. Die beiden Daten können gegenübergestellt werden um somit die Effizienz zu ermitteln. Hierfür muss ein adäquater Zeitraum bestimmt werden.

$$\text{uplift} = \frac{\text{Churn}_{\text{after}} - \text{Churn}_{\text{before}}}{\text{Churn}_{\text{before}}} \quad (10.3)$$

11 ML-Canvas

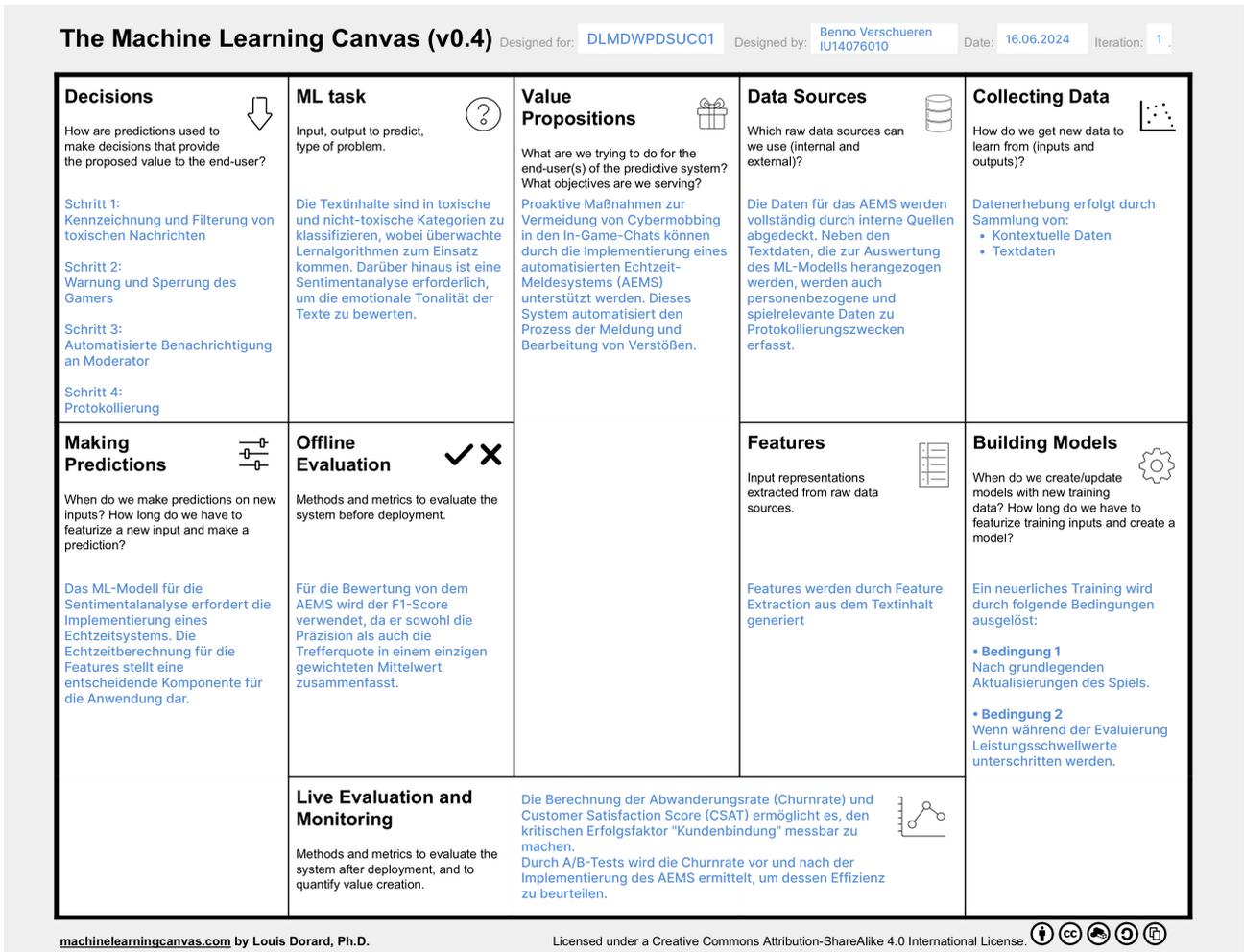


Abbildung 5: ML-Canvas für das AEMS (www.thecanvasrevolution.com (o.J.))

12 Pitchdeck

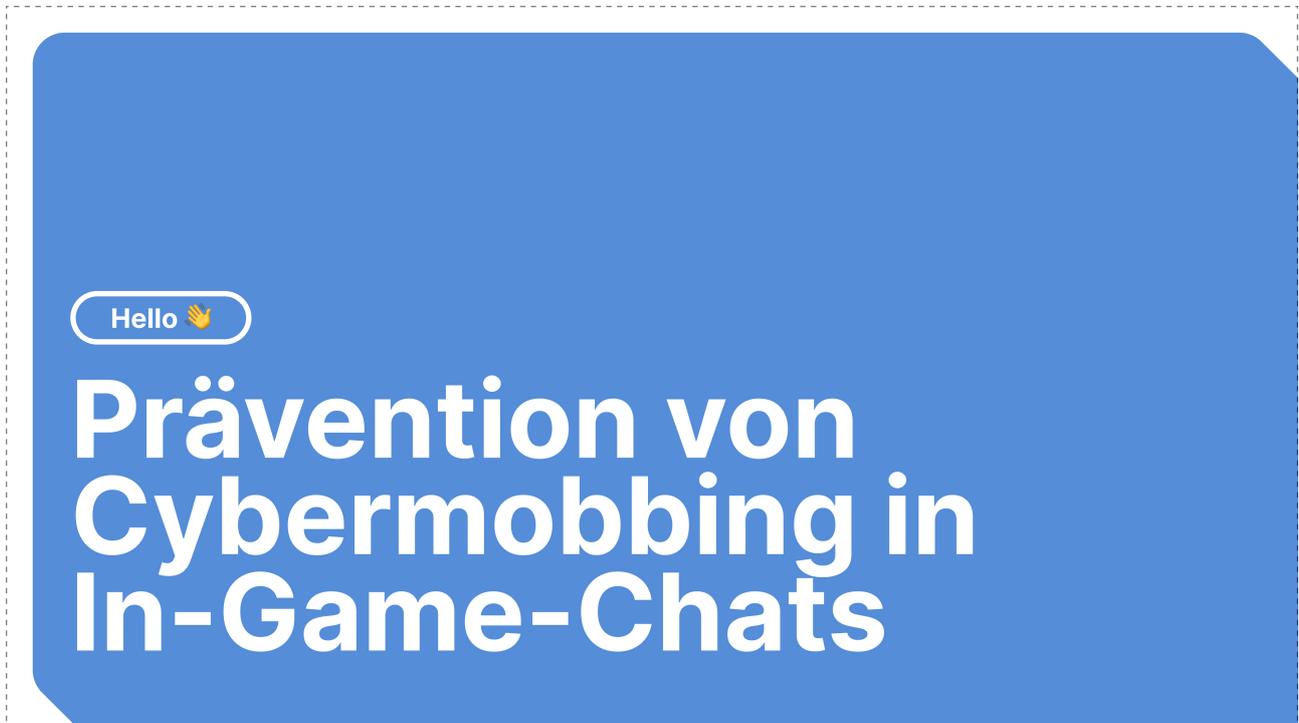


Abbildung 6: Pitchdeck Folie 1

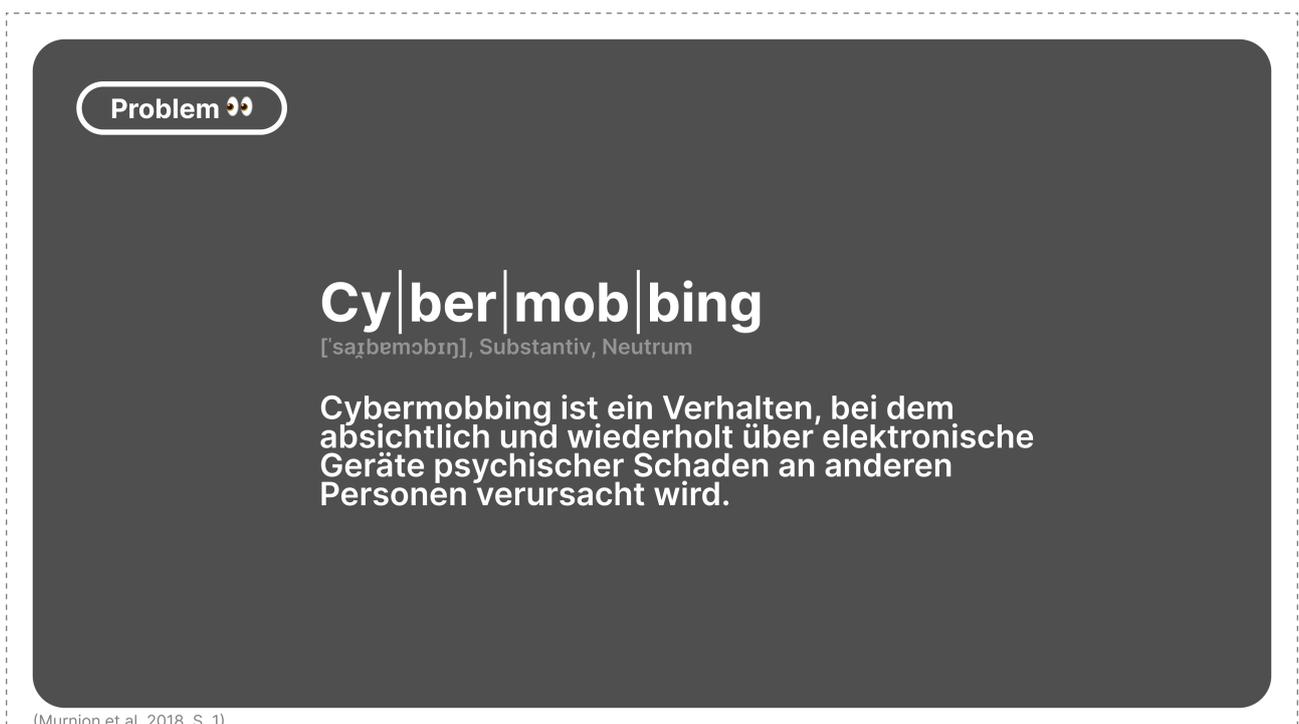


Abbildung 7: Pitchdeck Folie 2

Folgen von Cybermobbing 🙄

68%

der Online-Multiplayer-Gamer haben schwerwiegendere Misshandlungen erlebt.

**Beendigung
von Werbepartnerschaften**

**Schädigung
Markenruf**

**Hohe Abwanderungsrate
bei Gamer**

Verlust von Partnerschaften

28%

der Online-Multiplayer-Gamer meiden Spiele, in denen sie Belästigungen erfahren haben.

(getstream.io, 2022) (ADL, 2020)

Abbildung 8: Pitchdeck Folie 3

Ziel 🙌

Proaktive Maßnahmen zur Vermeidung von Cybermobbing erhöhen das Vertrauen und die Kundenbindung der Gamer sowie gegebenenfalls auch deren Eltern. Ein weiterer positiver Aspekt ist die medienwirksame und klare Positionierung von RagePlayer® gegen Hass im Netz.

Abbildung 9: Pitchdeck Folie 4

Lösung 🤔

Für RagePlayer® soll ein automatisiertes Echtzeit-Meldesystem (AEMS) implementiert werden, um Cybermobbing in den In-Game-Chats zu verhindern.

Basierend auf der Sentimentalanalyse wird ein KI-basiertes Klassifikationsmodell Nachrichten automatisch kennzeichnen und unleserlich darstellen, wenn sie gegen die Richtlinien von RagePlayer® verstoßen.

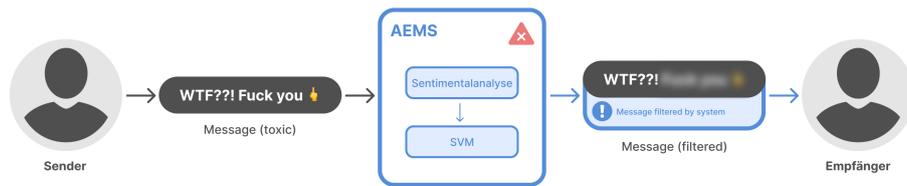


Abbildung 10: Pitchdeck Folie 5

Literaturverzeichnis

- ADL (2020):** *Free to play? hate, harassment and positive social experience in online games 2020.* URL: <https://www.adl.org/resources/report/free-play-hate-harassment-and-positive-social-experience-online-games-2020> [last access 06.01.2024]).
- Arkhipov, I. (2018):** *Extraktion und stimmungsanalyse von tweets bezüglich bestimmter schlüsselwörter.*
- Atoum, J. (2020):** *Cyberbullying detection through sentiment analysis.*
- Barkved, K. (2022):** *How to know if your machine learning model has good performance.* (URL: <https://www.obviously.ai/post/machine-learning-model-performance> [last access 08.05.2024]).
- Bernazzani Barron, S. (2023):** *How to calculate churn rate in 5 easy steps [definition + formula].* (URL: <https://blog.hubspot.com/service/what-is-churn-rate> [last access 21.05.2024]).
- Chen, Y./Farley, T./ Ye, N. (2004):** *QoS Requirements of Network Applications on the Internet.* Jg. 4.
- datascientest.com (2023):** *Semantische analyse: Definition und anwendungsbeispiele dieses konzepts.* (URL: <https://datascientest.com/de/semantische-analyse-definition-und-anwendungsbeispiele-dieses-konzepts> [last access 01.05.2024]).
- Dilmegani, C. (2024):** *Model retraining: Why and how to retrain ml models?* (URL: <https://research.aimultiple.com/model-retraining/> [last access 01.05.2024]).
- getstream.io (2022):** *The importance of moderating in-game chat.* (URL: <https://getstream.io/blog/video-game-chat-moderation/> [last access 11.05.2024]).
- Herring, S./Kutz, D./Paolillo, J./ Zelenkauskaitė, A. (2009):** *Fast Talking, Fast Shooting: Text Chat in an Online First-Person Game.* 1 – 10.
- Leiba, Y./Shirokopetleva, M./ Gruzdo, I. (2023):** *RESEARCH ON METHODS OF DETERMINING CUSTOMER LOYALTY AND ASSESSING THEIR LEVEL OF SATISFACTION.* 104–117.
- Medhat, W./Hassan, A./ Korashy, H. (2014):** *Sentiment analysis algorithms and applications: A survey.* Jg. 5, 1093–1113.
- Murnion, S./Buchanan, W. J./Smales, A./ Russell, G. (2018):** *Machine learning and semantic analysis of in-game chat for cyberbullying.* Jg. 76, 197–213.
- Parmenter, D. ():** *Key Performance Indicators - Developing, Implementing and Using Winning KPIs.* John Wiley & Sons.
- Petrov, C. (2024):** *50 alarming cyberbullying statistics to know in 2024.* (URL: <https://techjury.net/blog/cyberbullying-statistics/> [last access 06.01.2024]).

- Runge, J./Gao, P./Garcin, F./ Faltings, B. (2014):** *Churn prediction for high-value players in casual social games.* 1–8.
- Sanjay, N. (2022):** *Continuous training of ml models.* (URL: <https://medium.com/@nagasanjayvijayan/continuous-training-of-ml-models-7d8acaf44dda> [last access 01.05.2024]).
- Shahana, P./ Omman, B. (2015):** *Evaluation of Features on Sentimental Analysis.* Jg. 46, 1585–1592.
- Simba, K. (2024):** *Breaking down real-time machine learning systems.* (URL: <https://www.featureform.com/post/breaking-down-real-time-machine-learning-systems-mlops> [last access 08.05.2024]).
- Singh, N./ Jaiswal, U. (2023):** *A Detailed Sentiment Analysis Survey Based on Machine Learning Techniques.* Jg. 12, e29105.
- Siroker, D./ Koomen, P. (2013):** *A/B Testing: The Most Powerful Way to Turn Clicks Into Customers.* Wiley.
- Sokolova, M./Japkowicz, N./ Szpakowicz, S. (2006):** *Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation.* Jg. Vol. 4304, 1015–1021.
- Swati, B./ Rajesh, P. J. (2015):** *Sentiment Analysis - Methods, Applications and Challenges.*
- Thompson, J. J./Leung, B. H./Blair, M. R./ Taboada, M. (2017):** *Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model.* Jg. 137, 149–162.
- ULD (2010):** *Datenschutz in online-spielen.* (URL: <https://www.datenschutzzentrum.de/uploads/projekte/dos/leitfaden-dos.pdf> [last access 12.01.2024]).
- Wnkhade, M./Rao, A./ Kulkarni, C. (2022):** *A survey on sentiment analysis methods, applications, and challenges.* Jg. 55, 1–50.
- www.bundestag.de (2016):** *How to calculate churn rate in 5 easy steps [definition + formula].* (URL: <https://www.bundestag.de/resource/blob/483622/32b7fb4bb887873dabcbb2b085be08dc/WD-7-154-16-pdf-data.pdf> [last access 21.05.2024]).
- www.thecanvasrevolution.com (o.J.):** *The machine learning canvas.* (URL: <https://thecanvasrevolution.com/product/the-machine-learning-canvas> [last access 29.06.2024]).
- Zhang, H./Yang, J./Chen, C./Wang, S./Li, J./ Lu, M. (2023):** *Principles and Practices of Real-Time Feature Computing Platforms for ML.* Jg. 66, 77–78.

Eidesstattliche Erklärung

Hiermit versichere ich eidesstattlich, dass ich die vorliegende Arbeit eigenständig und ausschließlich unter Verwendung der im Quellen- und Literaturverzeichnis aufgeführten Werke angefertigt habe.

München, 16.06.2024

.....
Ort, Datum



.....
Unterschrift