Ganzheitlicher Optimierungsansatz zur Verbesserung von Instandhaltungsabläufen auf Basis von Social Media Inhalten

Stübinger, Lukas[1], Lomaeva Mariia[1], Vidal Migallon Irina[1],
Lehmann, Oliver[1]
[1]Siemens Mobility GmbH

Zusammenfassung

Um Digitale Services für den Fahrzeuginnenraum zu etablieren, ist es notwendig, die Servicekosten und Implementierungsaufwände gering zu halten. Hierfür wird im vorliegenden Ansatz auf Daten aus Social Media zurückgegriffen. Diese sind zwar frei verfügbar, liegen jedoch in einem unstrukturierten Format vor, sodass die automatisierte Analyse erschwert wird. Mithilfe des beschriebenen Ansatzes wird gezeigt, dass neuste Natural Language Processing Analysen in der Lage sind, diese Daten aufzubereiten und bezüglich definierter Themen für die Instandhaltung zu klassifizieren und auswertbar zu machen. Die Performance der verwendeten Techniken übertrifft die definierte Baseline und zeigt für die Instandhaltung taugliche Ergebnisse. Social Media kann also für die Gewinnung von Informationen für die Instandhaltung verwendet werden, um eine bessere Planbarkeit zu erzeugen und letztlich die Depotaufenthalte möglichst gering zu halten.

1 Einleitung

Um das Ziel von 100% Verfügbarkeit für die optimale Ausnutzung sowohl des Rollmaterials als auch der Infrastruktur zu erreichen, ist es notwendig, verschiedene Informationsquellen für den aktuellen Zustand des gesamten Eisenbahnsystems heranzuziehen. Dabei konzentriert sich die aktuelle Forschung verstärkt auf die Analyse und Auswertung von Sensorinformationen. Die Überwachung der kritischen Elemente des Rollmaterials (z. B. Drehgestell) und der Infrastruktur (z. B. Weiche) steht im Fokus der Bemühungen [15], um einen reibungsfreien Betrieb zu gewährleisten. Aus diesem Grund gewinnen die Digitalen Services zusehends mehr an Bedeutung, da durch diese Informationen über kritische Anlagen und Komponenten des Eisenbahnsystems für optimale Entscheidungen gesammelt werden können.

Die technischen Entwicklungsbestrebungen leisten ihren Beitrag für ein störungsfreies Fahrzeug. Für die Verkehrsbetriebe stehen aber auch ihre Passagiere und deren Komfort vermehrt im Kern ihrer Bemühungen. Damit wird zum einen eine hohe Systemverfügbarkeit erzielt und zum anderen die Attraktivität des schienengebundenen Verkehres für die Passagiere gestärkt. Um nicht nur die kritischen Elemente mithilfe von Digitalen Services abzudecken, ist es erforderlich, auch aus der Sicht der Passagiere tätig zu werden und deren Bedürfnisse bei der Entwicklung von Digitalen Services zu berücksichtigen. Dabei spielen geringe Servicekosten wie auch Implementierungsaufwände eine wichtige Rolle, da es sich vornehmlich um schlecht diagnostizierbare Komponenten im Fahrzeuginneren handelt.

In den letzten Jahren sind die analytischen Möglichkeiten für die Interpretation verschiedenster Daten und deren Anwendungen enorm angestiegen. Dabei wird vornehmlich auf Künstliche Intelligenz (KI) für die Auswertung von Sensordaten zurückgegriffen. Textanalysen jedoch werden bislang im Bahnbereich selten eingesetzt. Diese bieten ein großes Potential insbesondere hinsichtlich der Analyse von Social Media [8]. Die Verwendung von Social Media Analytics ist im Konsumentenbereich sehr stark verbreitet. Mittels Meinungs- und Stimmungsanalysen sowie weiteren Verfahren der Text- und Wortanalyse aggregieren Unternehmen beispielweise die Meinung der Verbraucher [18]. Diese Methoden sind bereits über mehrere Jahre erprobt [17]. Darüber hinaus wird eine Verdoppelung der aktiven Social Media Nutzer seit 2015 bis Ende 2021 erwartet, sodass die Bedeutung von Social Media in der Bevölkerung stark zunehmend ist [16]. Es liegt also zum einen ein sehr großes Potential in der Auswertung von Daten aus Social Media vor und zum anderen werden vor allem für Interieur-Komponenten des Fahrzeugs aktuell aus Prioritätsgründen wenig Bemühungen für deren Überwachung unternommen.

2 Problemstellung

Um für die Komponenten im Fahrzeuginneren Digitale Services zu etablieren, ist es erforderlich, geringe Servicekosten wie auch Implementierungsaufwände bei der Entwicklung zu berücksichtigen. Aufgrund der nicht wirtschaftlichen und aktuell mangelnden alternativen Überwachungsmöglichkeiten für den Innenraum des Fahrzeugs werden Social Media Posts für die Ableitung des Zustands in Betracht gezogen. Dabei treten abgesehen von den bekannten Herausforderungen der Textverarbeitung noch weitere Besonderheiten hinsichtlich der Verwendung für die Instandhaltung der Fahrzeuge auf.

Besonderheiten für die Verwendung in der Instandhaltung:

- Eindeutige Zuordnung zu Fahrzeugen

- Eindeutige Zuordnung zu Komponenten
- Bestimmung des Zustands

Um das vorgestellte Thema von den bisherigen Veröffentlichungen abzugrenzen und den aktuellen Stand der Forschung aufzuzeigen, wurden die folgenden Suchanfragen über Google Scholar am 06.07.2021 gestellt. Alle Suchanfragen wurden in der englischen Sprache durchgeführt. Die Ergebnisse der Suchanfragen sind in Tabelle 1 dargestellt. Bei der inhaltlichen Überprüfung ist festzustellen, dass der Input aus Social Media nicht in direkter Verbindung zu Instandhaltungsaktivitäten gebracht wird. Zwar wird beispielsweise darauf eingegangen, dass Social Media für die Kommunikation mit dem Passagier über dessen Erfahrungen während der Reise verwendet wird, aber es wird keine direkte Weiterverwendung von Social Data als Input für die Planung der Instandhaltungsaktivitäten thematisiert [9].

Tabelle 1: Durchgeführte Suchanfragen und die Anzahl der Ergebnisse.

Themengebiet	Suchbegriffe	Ergebnisse
Social Media im Kontext von Instandhaltung im Allgemeinen	social media; maintenance	ca. 2 830 000
Social Media im Kontext von Instandhaltung in der Luftfahrt	social media; maintenance; aviation	ca. 63 500
Social Media im Kontext von Instandhaltung im Schienenverkehr	social media; maintenance; railway	ca. 103 000

3 Zielstellung

Aufgrund des großen Potentials von Social Media für die Gewinnung von instandhaltungsrelevanten Informationen wurde ein Ansatz entwickelt und dessen Funktionsweise und Performance überprüft. Die gewonnenen Informationen werden als Planungsgrundlage der Instandhaltungsabläufe im Depot herangezogen. Bei einem geeigneten Instandhaltungsprojekt wurden für diese Untersuchung über einen Zeitraum von über zwei Jahren Twitter-Nachrichten aufgezeichnet und hinsichtlich ihrer Verwendbarkeit für die Instandhaltung bewertet. Dabei kommen für die Analyse der Nachrichten sogenannte Natural Language Processing Algorithmen zum Einsatz. Diese Algorithmen ermöglichen es, die Informationen aus den Nachrichten zu klassifizieren, zu analysieren, aufzuarbeiten und zusammenzufassen, sodass deren Inhalt für die Optimierung der Instandhaltungsabläufe verwendet werden kann. Das Ziel ist es, Zustände von Komponenten im Fahrzeuginnenraum abzuleiten, welche die Passagiererfahrung beeinflussen. Diese Informationen über den Zustand werden anschließend für die Instandhaltungsplanung aufbereitet und zur Verfügung gestellt.

4 Vorgehen

4.1 Datensammlung

Die verwendeten Analysemethoden basieren auf überwachten Lernmethoden für Klassifikationsprobleme [5], welche annotierte Trainingsdatensätze benötigen. Dafür wurde mittels Datenabfrage nach Stichwörtern, Mentionings und Hashtags des untersuchten Bahnbetreibers, bei zwei Social Media Aggregatoren eine Datensammlung durchgeführt. Im Zeitraum 01.01.2019 bis 31.05.2021 wurden mehr als 270 000 Tweets abgefragt, wobei Retweets herausgefiltert wurden. Bei dem Betreiber handelt es sich um ein Nahverkehrsunternehmen, bei dem in der Spitze zu Nicht-Pandemiezeiten bis zu 24 Züge pro Stunde auf den Hauptverkehrsstrecken fahren.

Zuerst wurden nur englischsprachige Tweets herangezogen und anschließend mittels regulärer Begriffe basierend auf einer umfangreichen Liste von Stichwörtern gefiltert. Die benutze Stichwortliste wurde aus einem umfangreichen Wortschatz gebildet, welcher aus vorhergehenden Twitter-Sprachanalysen, betriebsinternen Kommunikationen zur Instandhaltung von Rollmaterial als auch Instandhaltungsapplikationen im Endkundeneinsatz gewonnen wurde. Hierbei wurde darauf Wert gelegt, keine möglicherweise instandhaltungsrelevanten Tweets auszuschließen. Dies wurde anschließend in einem manuellen Arbeitsschritt überprüft.

Die beschriebene Vorfilterung ergibt eine Menge von ca. 21 000 Tweets, die nachfolgend annotiert wurden und von denen ca. 4 000 Tweets als instandhaltungsrelevant identifiziert wurden. Nichtrelevante Tweets, also Tweets, die keine Informationen zu Rollmaterial und/oder instandhaltungsrelevanten Themen beinhalteten, behandelten Themen wie z. B. COVID-19, Ticketreservierungen oder allgemeine Kommentare zu Betriebsabläufen.

Die Liste der instandhaltungsrelevanten Themen wurden aus den Produktgruppen des betrachteten Rollmaterials gewonnen und rekombiniert, sodass Trainingsdaten mit ausbalancierten Klassen entstehen. In Abbildung 1 ist sowohl das Ausbalancieren der Themen als auch die Anzahl der Tweets in der jeweiligen Klasse dargestellt. Die schlussendlich genutzten Kennzeichnungen der instandhaltungsrelevanten Themen sind:

- *wifi* - Zugang zum Internet
- *hvac* – Heizung, Lüftung, Klimatechnik

- *seats/tables* –Fahrgastsitze und -tische
- *announcements* – Ansagen und Anzeigen zum Fahrablauf
- *vandalism* – Schäden und Auswirkungen durch Vandalismus (z. B. Graffiti)
- *toilets* – Sanitäre Anlagen
- *interior* – Fahrgastraum
- *doors* – Fahrgasttüren
- *brakes/noise* – Verkehrslärm und Bremsverhalten

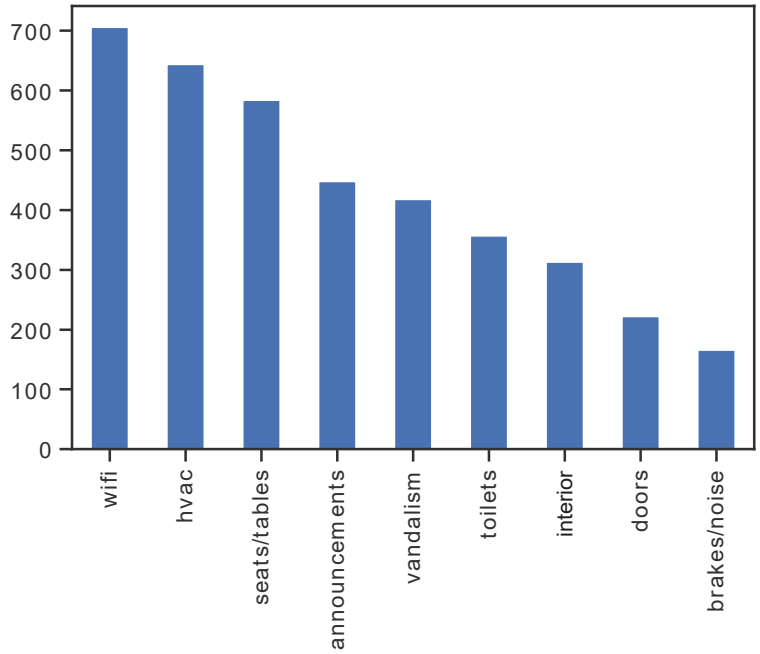


Abbildung 1*: Anzahl instandhaltungsrelevanter Tweets im betrachteten Zeitraum.*

Die Daten wurden von mehreren Personen annotiert und kategorisiert, denen nichtüberschneidende Gruppen von Tweets zugeteilt wurden. Idealerweise sollte jedes Tweet-Label durch die Mittelung mehrerer Urteile von Muttersprachlern bestimmt werden, um eine hohe Annotierungsgüte zu erreichen. Dies ist von besonderem Interesse, wenn komplexe Sprache, Sarkasmus oder Inhalte identifiziert und beurteilt werden, die spezifisch für die untersuchten Regionen sind und manchmal nur von ortskundigen Passagieren verstanden werden können. In der durchgeführten Analyse wurde die Mittelwertmethode durch Konsensbildung für Tweets, welche nicht eindeutig zu oben genannten Kennzeichnungen zugeordnet werden konnten, angewandt. Dies reduzierte den Bearbeitungsaufwand um ca. 60% ohne die Annotierungsgüte ausschlaggebend zu verringern.

4.2 Datenanalyse

Die durchgeführte Datenanalyse gibt Einblicke in die gesammelten Texte und gibt Aufschluss über die Verteilung der Tweets in verschiedenen Zeiträumen. Wie in Abbildung 2 ersichtlich wird, bleibt die Verteilung der Tweets während kurzer Zeithorizonte (Größenordnung eine Woche) über den untersuchten Zeitraum annähernd konstant. Die größere Anzahl von Tweets im Jahr 2019 im Vergleich zu den Niveaus in den Jahren 2020 und 2021 spiegelt die geringere Anzahl von Passagieren aufgrund der erlassenen Richtlinien zur Eindämmung der COVID-19 Pandemie wider. Insbesondere die Ausgangssperre im Frühjahr 2020 zeigt sich deutlich in der Analyse sowie eine zweite Welle im Herbst 2020.

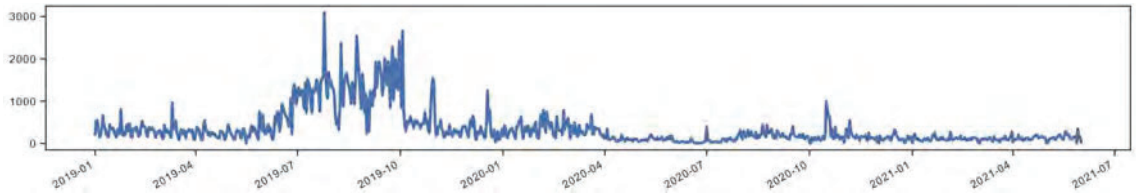


Abbildung 2: Anzahl Tweets über den Betrachtungszeitraum.

In Abbildung 3 wird die Tweet-Verteilung für ausgewählte, instandhaltungsrelevante Themen nach Tageszeit gezeigt. Die Anzahl der Tweets korreliert mit den Hauptverkehrszeiten, woraus sich schließen lässt, dass Fahrgäste ihre Erfahrungen ohne oder nur mit geringer Verzögerung berichten. Ferner ist ersichtlich, dass Fahrgäste vermehrt in der zweiten Tageshälfte Nachrichten posten, die sanitäre Anlagen betreffen.

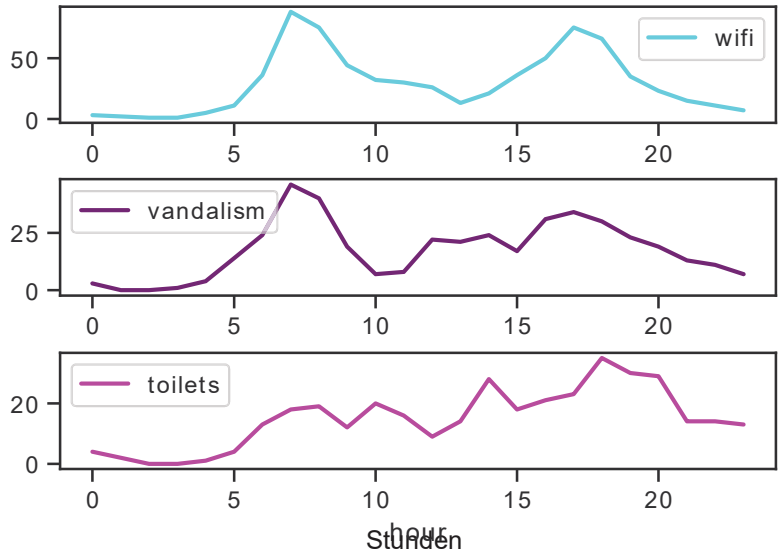


Abbildung 3: Häufigkeit der Tweets für ausgewählte Themen nach Tageszeit.

In Abbildung 4 wird die Häufigkeit der Erwähnungen instandhaltungsrelevanter Themen nach Wochentag dargestellt. Obwohl Fahrgäste an Wochenenden grundsätzlich weniger posten (siehe Samstag & Sonntag für *wifi*), äußern sie am Wochenende häufiger Probleme mit den Toiletten, was möglicherweise auf ein Reinigungsdefizit an diesen Tagen hinweist.

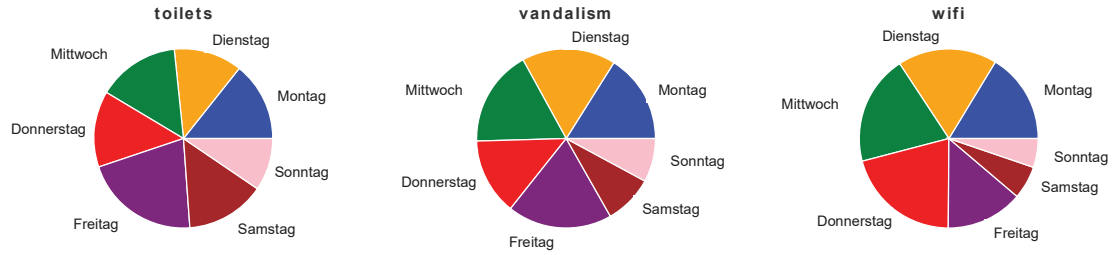


Abbildung 4: Häufigkeit der Tweets für ausgewählte Themen nach Wochentag.

In Abbildung 5 wird die normalisierte Tweet-Frequenz im Zeitverlauf gezeigt, wobei 0 und 1 keine Tweets bzw. maximale Anzahl von Tweets pro Monat darstellen. Hier zeigt sich, dass sich die Bedenken der Passagiere im Laufe der Zeit verschieben. Während der COVID-19 Pandemiezeiten wurden vermehrt Tweets zu den Themen Fahrgastsitze und -tische (*seats/tables*) beobachtet, als Fragen der sozialen Distanzierung (Abstand der Sitzplätze), Hygiene (Reinigung der Kontaktflächen) und Belüftung wichtiger wurden. Einige Themen sind jedoch gleichbleibend beliebt, wie z. B. Internet (*wifi*), Toiletten (*toilets*) oder Vandalismus (*vandalism*).

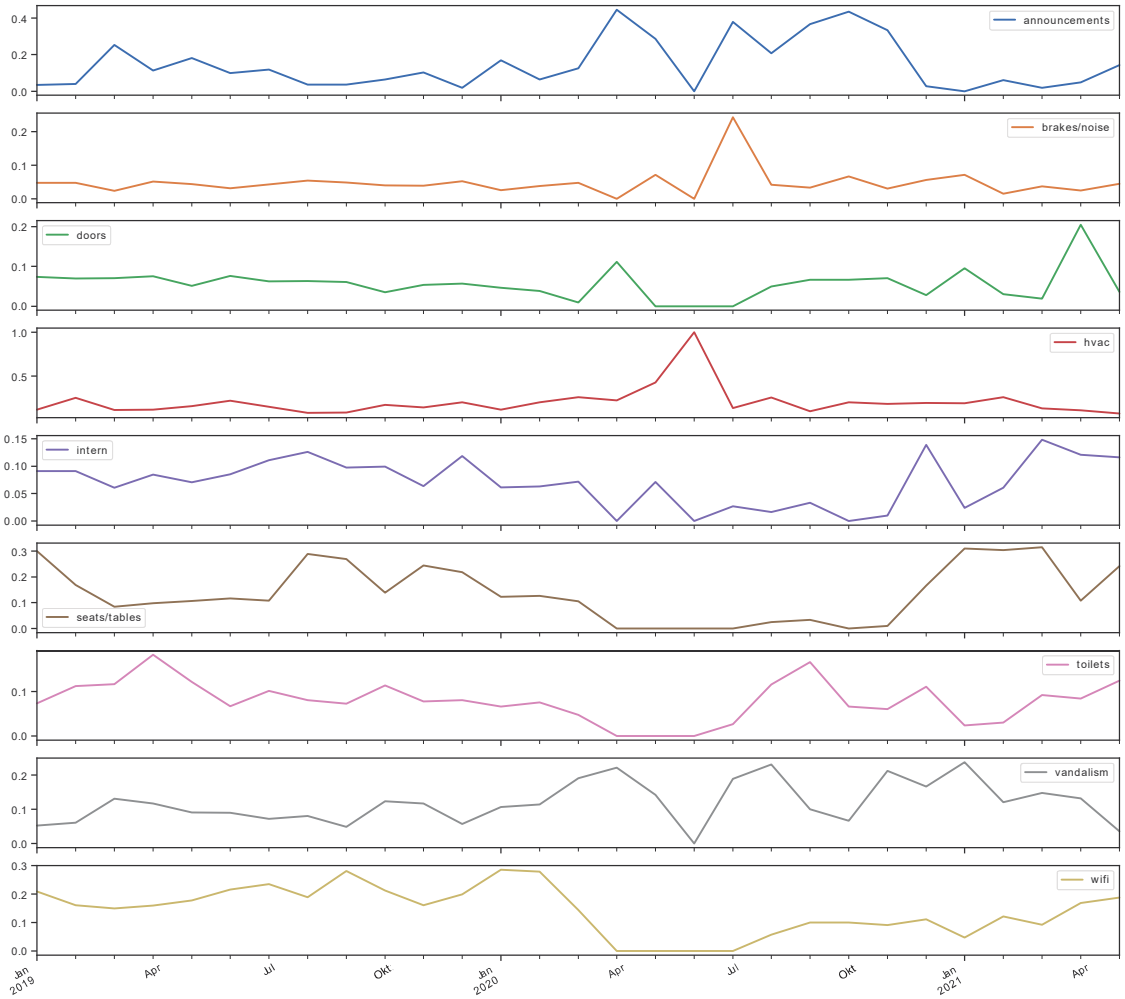


Abbildung 5: Normalisierte, monatliche Tweet-Frequenz über den Betrachtungszeitraum

Im Generellen zeigt die Analyse, dass Fahrgäste den Fahrgastkomfort, Service und Ansagen zum Fahrbetrieb am häufigsten kommentieren. Dazu werden von den Kunden oftmals noch Vorschläge zur Verbesserung der vorgenannten Aspekte gepostet.

4.3 Tweet Klassifikation

Für den Ansatz wurde ein Praxisbeispiel ausgesucht, bei welchem ausreichend Twitter-Nachrichten vorhanden sind. Einige davon sind jedoch selten relevant für den Betrieb des Rollmaterials und dessen Instandhaltung. Um den Inhalt zu identifizieren, welcher instandhaltungsrelevante Themen anspricht, werden alle Tweets in folgendem, zweistufigen Prozess gefiltert:

1. Klassifizierung nach Relevanz (relevanter Tweet vs. nicht relevant); und anschließend für relevante Tweets,
2. Bestimmung des Themas mittels Multiklassen-Klassifizierung

4.3.1 Baseline

In der ersten Phase der Entwicklung neuer KI-basierter Analysen ist es vorteilhaft eine Vergleichsbasis zu erstellen, um optimierte Methoden bewerten zu können. Das Basis-Klassifikationsmodell basiert auf traditionellen, maschinellen Lernalgorithmen. Für die oben genannten Prozessstufen wurden SVM (Support Vector Machines) [14], Random Forest [4], logistische Regression [7] und Entscheidungsbäume [3] getestet. Diese Algorithmen klassifizieren die Tweets angemessen, sind aber nicht in der Lage die komplexe Struktur der Sprache zu erfassen, die in Twitter-Nachrichten oft benutzt wird.

Die erste Klassifikationsstufe ist eine binäre Auswahl, bei der Tweets als relevant oder nicht relevant klassifiziert werden. Die beste Performance für den Relevanztest wurde mit dem Support Vector Classifier (SVC)-Algorithmus erreicht. Hierbei werden die Tweets mittels TF-IDF [11] Vectorizer in Vektoren umgewandelt, der jedem N-Gramm von Fragmenten mit eins bis drei Wörtern Länge entsprechend ihrer Bedeutung im Text eine Punktzahl zuweist. Die vektorisierte Repräsentation der Tweets wird dann dem SVC zugeführt. Die Performanz von SVC bei den Testdaten ist in Tabelle 2 zu sehen.

Der zweite Klassifikationsschritt bestimmt das Instandhaltungsthema für relevante Tweets. Hierbei wird zwischen den neun o.g. verschiedenen Themen in einer Multiklassen-Klassifikation unterschieden. Als Basislinienalgorithmus wurde ein Random Forest ausgewählt und trainiert. Dieses Modell besteht aus einem Ensemble von Entscheidungsbäumen, die zuerst jeweils eine Klasse (Thema) pro Tweet bestimmen, aus denen anschließend per Mehrheitsentscheid das finale Thema gewählt wird. Einer der Hauptvorteile von Random Forests ist ihre Interpretierbarkeit. Es ist relativ einfach zu beurteilen und zu visualisieren, wie und warum der Algorithmus eine bestimmte Entscheidung getroffen hat.

Die Herausforderung für den zweiten Klassifikationsschritt liegt in der größeren Anzahl von Themen (9 statt 2 Klassen) und dass einem Tweet mehr als ein Thema zugeordnet werden kann. Wie im ersten Schritt werden die Tweets mit TF-IDF vektorisiert und anschließend durch den Random Forest klassifiziert. Die Gesamtklassifizierungsmetrik, der macro F_1-Maß [2], ist mit 0,7 zwar besser als eine rein zufällige Themenzuordnung (0,11), aber stellt noch kein zuverlässiges Ergebnis dar.

4.3.2 Klassifikation mittels Transformer

Ein Transformer [19], im Kontext des maschinellen Lernens zur Verarbeitung natürlicher Sprache, ist ein Deep-Learning-Modell [10], der die Bedeutung aller Bestandteile von

Eingabedaten (hier Twitter-Nachrichten) bewertet. Transformer benutzen einen Aufmerksamkeitsmechanismus, der den Kontext von zu untersuchenden Satzfragmente identifiziert und damit deren Bedeutung bestimmt. Dies kann an folgendem Beispiel veranschaulicht werden:

Didn't get a window seat on the train ride today because it was so crowded!

Didn't get a window seat on the train ride today because it was all torn!

Da beide Sätze fast ausschließlich aus den gleichen Wörtern gebildet werden und die oben beschriebenen Basisalgorithmen Entscheidungen aufgrund von Metriken verwenden, die auf der Worthäufigkeit in Texten basieren, würden die Basisalgorithmen diese Beispielsätze als annähernd äquivalent ansehen. Der Unterschied jedoch, welcher auf einen Mangel hinweist („Zug voll“ vs. „zerrissener Sitz“), würde nicht erkannt werden. Vor allem in solch ambivalenten sprachlichen Situationen können Transformer ihre Stärke beweisen und erkennen, dass sich die Aussage des ersten Satzes auf einen Zug bezieht, während der zweite ein Sitzplatzproblem behandelt.

Der manuell annotierte Datensatz wurde für das Training aller Transformer benutzt. Um die negativen Auswirkungen des stark unausgeglichenen Datensatzes von etwa 4 000 relevanten Tweets und 17 000 irrelevanten Tweets entgegenzuwirken, wurde der endgültige Datensatz, bestehend aus 7 700 Tweets, durch die Auswahl aller relevanten Tweets und der gleichen Anzahl zufällig ausgewählter irrelevanten Tweets gebildet. Für beide Klassifikationstrainings (Relevanz- und Themenphase) wurde der Datensatz in Trainings-, Validierungs- und Testdatensatz aufgeteilt. 80% des Datensatzes wurden für das Training und die restlichen 20% für Validierung und Tests verwendet.

Die besten Ergebnisse bei der Themenklassifizierung (Phase 2) wurden mit „Bidirectional Encoder Representation from Transformers“ (BERT) [6] erzielt. Für den Relevanzklassifizierungsschritt (Phase 1) wurde schließlich RoBERTa gewählt, welcher den „Robust BERT Approach“ [13], der BERT übertraf. Die Kreuzentropie-Funktion [20] wurde verwendet, um die Vorhersagekraft der Modelle (*loss*) zu messen. Für den Optimierungsalgorithmus des Modelltrainings, der die Gewichte (oder Konfiguration) des Transformer-Modells bei jeder Iteration entsprechend dem *loss* kontinuierlich aktualisiert, wurde der Adam-Optimierer [12] sowohl für die Relevanz- als auch für die Themenklassifizierung ausgewählt.

Die Güte der Modelle wurde durch Hyperparameter-Tuning optimiert, einschließlich der-Lernrate, der Batchgröße und der Anzahl der Trainingsepochen [10]. Das Training wurde auf GPUs mit Hilfe des FLAIR-Bibliothek [1], das State-of-the-Art-Lösungen für Deep

Learning bereitstellt, durchgeführt. Alle trainierten Transformer wurden auf dem Testdatensatz mit den folgenden Metriken bewertet: Genauigkeit, F1-Score, Präzision und Erinnerung [2].

Für Phase 2, ebenfalls in Tabelle 2 dargestellt, wurde der gewichtete Mittelwert des F1-Wertes aller Klassen als Kennzahl der Güte der Mehrklassen-Klassifizierung benutzt. Der höchstmögliche Wert ist 1, was perfekte Genauigkeit und Trefferquote anzeigt, und der niedrigste ist 0.

Tabelle 2: Güte der Algorithmen für beide Klassifikationsphasen.

	Baseline	Transformer
Phase 1: Relevanz (macro F1-Maß)	0.84	0.937
Phase 2: Topics (gewichteter F1-Mittelwert)	0.70	0.955

Im Generellen ist die in sozialen Medien benutzte Sprache schwer maschinell zu analysieren, da sie häufig komplex ist. Es treten u. a. viele subtilen Redenwendungen, Slang, gewollten Stilblüten und Rechtschreibfehlern auf. Sogar für Muttersprachler ist die benutze Sprache zum Teil schwer verständlich. Einige Tweets weisen eine gewisse Mehrdeutigkeit auf, die sogar menschliche Leser herausfordert, was in einigen Fällen zu Unstimmigkeiten bei der Themenklassifikation führen kann, was wiederum das Training der Transformer erschwert. Trotz der nicht perfekten Annotation sind Transformer jedoch oft in der Lage, die wahre Themenklassifikation selbst zu erlernen, wenn sie mit genügend Daten trainiert werden. Der Schlüssel zu den weiteren Verbesserungen des eingesetzten Klassifikationsmodells liegt daher in noch größeren, beschrifteten Datensätzen mit einer breiteren Auswahl instandhaltungsrelevanter Tweets. Auch hier bestätigt sich wieder, dass alle Modelle des überwachten, maschinellen Lernens, selbst die fortschrittlichen wie Transformers, nur so gut sind wie die Daten, durch welche sie trainiert wurden.

4.4 Nutzen im Anwendungsfall

An einem ausgewählten und anonymisierten Praxisbeispiel wird gezeigt, wie der Ansatz neuen Input für die Instandhaltungsplanung generiert und wie diese prozessual in den Ablauf einfließen. Über den ausgewählten Zeitraum wurden im Praxisbeispiel Daten gesammelt und analysiert. Für die Anwendung ist es wichtig die Abläufe im Depot besser zu verstehen und genau zu betrachten, zu welchem Zeitpunkt der neue Input zum Tragen

kommt. Aktuell werden Fehler im Innenraum bei Routineinspektionen nach der Einfahrt des Zugs im Depot festgestellt. Um im Vergleich dazu Zeit einzusparen, die Qualität der Tätigkeiten zu steigern und die Passagiererfahrung zu verbessern, ist es erforderlich, dass die Information vor Einfahrt des Zugs vorliegt oder Fehler erkannt werden, die nicht bei der Routineinspektion festgestellt werden. Deswegen ist es notwendig, dass die analysierten Informationen in der Arbeitsvorbereitung und in der Ablaufplanung zu Beginn einer Schicht vorliegen. Ist dies der Fall, so kann die Arbeitsvorbereitung noch vor Einfahrt eines Zugs im Depot feststellen, ob und welche Instandhaltungsmaßnahmen im Innenraum des Zuges anstehen. Damit können diese Maßnahmen vorbereitet werden, um die Zugaufenthalte im Depot möglichst kurz zu halten. Dies führt letzten Endes zum einen zu einer besseren Planbarkeit und zum anderen zu einer besseren Planungsqualität. Abgesehen von der Verbesserung der Planbarkeit im Depot können ebenfalls unerkannte Fehler entdeckt werden. Beispielsweise wurde über Social Media eine nicht funktionierende Steckdose erkannt, welche durch eine visuelle Inspektion allein nicht entdeckt worden wäre. Bei der Durchsicht des verwendeten Instandhaltungssystems war dieser Fehler ebenfalls nicht aufzufinden. Da die Verbesserung der Analysen noch fortschreitet, ist noch keine abschließende Verifizierung aller erkannten Fehler mit dem Instandhaltungssystem durchgeführt worden. Es hat sich aber gezeigt, dass der vorgestellte Ansatz, sofern über die Defizite im Innenraum auf Social Media berichtet wird, verwendet werden kann, um diverse Komponenten zu überwachen.

5 Zusammenfassung und Ausblick

Der vorgestellte Ansatz zeigt, dass mittels Social Media-Daten eine weitere Datenquellen für die Beurteilung des Zustandes der Komponenten im Fahrzeuginnenraum herangezogen werden kann. Im Praxisbeispiel wurde die prozessuale Einbindung in die Instandhaltungsabläufe erklärt und ein erstes Beispiel für den Nutzen aufgezeigt.

Beim vorliegenden Ansatz wurden neuste Natural Language Processing Algorithmen zur Anwendung gebracht. Im Vergleich zu der beschriebenen Baseline weisen diese eine gute Performance auf. Bei der Datenaufnahme und -Analyse gibt es jedoch immer noch Verbesserungspotential. Um sowohl die Relevanz- als auch die Themenklassifizierung zu optimieren, ist es notwendig, die Datenerhebung und -annotation fortzusetzen. Auch wenn dies für die meisten datenbasierten Analysemethoden gilt, wurde das Problem durch die während der Coronavirus-Pandemie verzerrten Daten verschärft, da der untersuchte Nahverkehrsanbieter für den größten Teil des Zeitraums von März 2020 bis Mai 2021 deutlich reduziert Fahrgastzahlen meldete.

Auch wenn es aufschlussreich war, ein so außergewöhnliches Ereignis zu untersuchen, werden die Social Media-Daten in der Zeit nach der Pandemie sicherlich nicht der Datenverteilung von 2020-2021 folgen. Diese Tatsache sollte berücksichtigt werden, wenn entwickelte Methoden eingesetzt werden.

Literatur

[1] AKBIK, Alan, et al. FLAIR: *An easy-to-use framework for state-of-the-art NLP*. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations). 2019. S. 54-59.

[2] BONNIN, Rodolfo.: *Machine Learning for Developers: Uplift your regular applications with the power of statistics, analytics, and machine learning*. In Packt Publishing Ltd, 2017.

Bonnin, Rodolfo. Machine Learning for Developers: Uplift your regular applications with the power of statistics, analytics, and machine learning. In Packt Publishing Ltd, 2017.

[3] BREIMAN, Leo, et al.: *Classification and regression trees*. In Routledge, 2017.

[4] BREIMAN, Leo.: *Random forests.* In Machine learning 45.1 (2001): 5-32.

[5] BURKOV, Andriy.: *The hundred-page machine learning book*. Vol. 1. Canada: Andriy Burkov, 2019.

[6] DEVLIN, Jacob, et al.: *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805 (2018).

[7] FRANKLIN, James.: *The elements of statistical learning: data mining, inference and prediction.* In The Mathematical Intelligencer 27.2 (2005): 83-85.

[8] GHOFRANI Faeze, et al: *Recent applications of big data analytics in railway transportation systems: A survey*. In Transportation Research Part C: Emerging Technologies, 2018

[9] GOLIGHTLY, David and Houghton, R.J.: *Social media as a tool to understand behaviour on the railways.* In Innovative applications of big data in the railway industry, IGI Global: Hershey, 224–239, 2018

[10] GOODFELLOW, Ian, Yoshua Bengio, and Aaron Courville.: *Deep learning*. In MIT press, 2016.

[11] JURAFSKY, Daniel, and James H. Martin.: *Speech and Language Processing: An Introduction to Natural Language Processing*. In Computational Linguistics, and Speech Recognition.

[12] KINGMA, Diederik P., and Jimmy Ba.: *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980 (2014).

[13] LIU, Yinhan, et al.: *Roberta: A robustly optimized bert pretraining approach.* arXiv preprint arXiv:1907.11692 (2019).

[14] NOBLE, William S.: *What is a support vector machine?.* In Nature biotechnology 24.12 (2006): 1565-1567.

[15] PACHL, Jörn: *Systemtechnik des Schienenverkehrs: Bahnbetrieb planen, steuern und sichern*. 8., überarbeitete und erweiterte Auflage. Wiesbaden: Springer Vieweg, 2016

[16] STATISTA: *Anzahl der aktiven Social-Media-Nutzer weltweit in den Jahren 2015 bis 2021*. https://de.statista.com/statistik/daten/studie/739881/umfrage/monatlich-aktive-social-media-nutzer-weltweit/, abgerufen am 24.06.2021

[17] STIEGLITZ, Stefan, Dang-Xuan, L., Bruns, A. et al: *Social Media Analytics.* In Wirtschaftsinformatik 56, 101–109, 2014

[18] STIEGLITZ, Stefan: *Social Media Analytics für Unternehmen und Behörden*. In Reuter C. (eds) Sicherheitskritische Mensch-Computer-Interaktion. Springer Vieweg, Wiesbaden, 2018

[19] VASWANI, Ashish, et al.: *Attention is all you need.* In Advances in neural information processing systems. 2017.

[20] ZAFAR, Iffat, et al.: *Hands-on convolutional neural networks with TensorFlow: Solve computer vision problems with modeling in TensorFlow and Python*. In Packt Publishing Ltd, 2018.

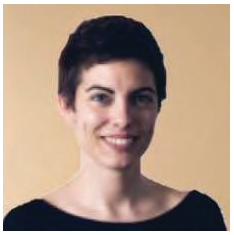
Autoren

Stübinger, Lukas

Studium des Maschinenbaus an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Seit 2018 bei der Siemens Mobility GmbH im Bereich Customer Services als Projektmanager Service Engineering Rail Infrastructure.

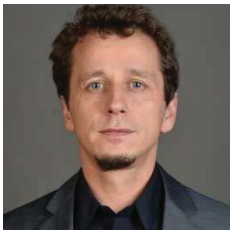
Lomaeva, Mariia

Studium der Computerlinguistik und Data Science an der Universität Potsdam und Hasso-Plattner-Institut für Digital Engineering. Seit 2020 bei der Siemens Mobility GmbH im Bereich Natural Language Processing."



Vidal Migallon, Irina

Studium der Medizintechnik und der Intelligente Systeme an der Technischen Universität von Madrid. Seit 2017 bei der Siemens Mobility GmbH im Bereich Digitalisation, Technology & IT und seit 2020 als Senior Key Expert für AI.



Lehmann, Oliver

Promotion an der Northeastern University, Boston, USA, im Bereich Complex System Modelling. Seit 2016 bei Siemens AG als Senior Scientist & Technology Consultant, mit Fokus auf der Entwicklung von analytikgetriebenen Entscheidungslösungen im Gesundheitswesen und im Bereich Mobilität.