

PROGNOSE DER RESTLEBENSDAUER EINES TRIEBWERKS MIT HILFE VON MASCHINELLEN LERNALGORITHMEN

H. Simon, U. Klingauf

Technische Universität Darmstadt, Institut für Flugsysteme und Regelungstechnik
Otto-Berndt-Straße 2, 64289 Darmstadt, Deutschland

Kurzfassung

Im vorliegenden Vorgehen wird ein datenbasierter Analyseansatz zur Diagnose des Gesundheitszustands eines Triebwerks und im Weiteren zur Prognose der Restlebensdauer vorgestellt. Die Datenanalyse basiert auf einer Reihe von run-to-failure Triebwerkssimulationsdaten, die vom NASA Prognostics Center of Excellence zu Forschungszwecken zur Verfügung gestellt werden. Der Datensatz wird speziell zu Forschungszwecken erstellt und ermöglicht den Vergleich von verschiedenen Modellansätzen. Auf den Datensatz werden unterschiedliche Algorithmen aus dem Bereich des Maschinellen Lernens, wie Klassifikations- sowie Regressions- und Clusteranalysen, angewendet. Als Zielgröße dient zum einen die Klassifikation von verschiedenen Gesundheitszuständen wie auch die Prognose der Restlebensdauer des Triebwerks.

Im Rahmen des Vortrages werden geeigneter Filter- und Glättungsalgorithmen zur Datenvorverarbeitung auf den Datensatz vorgestellt. Verschiedene Konzepte zur Diagnose des Gesundheitsstatus und zur Prognose der Restlebensdauer werden diskutiert. Zudem werden die verschiedenen Ansätze anhand geeigneter Performanzkriterien verglichen, um die automatisierte Klassifikation der Daten bewerten zu können.

Ausgehend von den vorhandenen Datensätzen wird der Vortrag Anforderungen an ein auf Realdaten basierendes Diagnosesystem des Aircraft Health Status diskutieren. Dies ist ebenfalls Untersuchungsinhalt im Projekt *INDI* (Intelligente Datennutzung in der Instandhaltung) im aktuellen Luftfahrtforschungsprogramm des Bundes.

1. EINFÜHRUNG UND RELEVANZ

Die Prognose von Bauteilversagen gewinnt erheblich an Aufmerksamkeit im Zusammenhang mit Aircraft Condition Monitoring Systemen, wie sie von Fluggesellschaften, Wartungsbetrieben und zunehmend auch von Triebwerksherstellern verwendet werden. Dadurch sollen korrektive Wartungsmaßnahmen reduziert und der Weg zur prädiktiven Instandhaltung geebnet werden. Hierfür ist eine fundierte Diagnose des Systemzustandes unabdingbar. Vorhersagen von Bauteilversagen mit einer gewissen Sicherheit geben zu können, ist aufgrund der Unsicherheiten des Verschleißprozesses eine Herausforderung. Sowohl datenbasierte als auch physikalische Modellansätze bieten Vorteile abhängig vom speziellen Anwendungsfall. [4] Der Mangel an vorhandenen *run-to-failure* Datensätzen ist eine große Einschränkung in der Entwicklung von datenbasierten Analysemethoden. In vielen Fällen sind Daten mit Verschleißanzeichen vorhanden, jedoch liegen sehr selten Daten bis zum kompletten Versagen der Komponente vor (siehe [2]).

In der vorliegenden Arbeit wird ein Analyseframework zur Diagnose und Prognose der Restlebensdauer (Remaining Useful Life, RUL) basierend auf simulierten Triebwerksdaten vorgestellt. Durch eine statistische Voruntersuchung soll dabei zunächst das Datenverständnis erhöht werden und eine fundierte Basis für das nachfolgende Analyseframework geschaffen werden. Aufbauend darauf wird das Datenanalyseframework mithilfe Methoden aus dem Bereich des maschinellen Lernens die Diagnose und Prognose des Systemzustands vornehmen.

Diese Arbeit kann damit als Grundlage für weitere Untersuchungen im Rahmen vom Forschungsprojekt *INDI* (Intelligente Datennutzung in der Instandhaltung) im aktuellen Luftfahrtforschungsprogramm des Bundes dienen. Im Rahmen des Projekts sollen Wartungsdaten, welche durch den Projektpartner Lufthansa Technik zur Verfügung gestellt werden sollen, für intelligente Datenanalysen genutzt werden. Dabei stehen Datenfusionsstrategien zur Unterstützung in der Flugzeugwartung im Vordergrund.

Im Folgenden wird in Kapitel 2 zunächst das angestrebte Analyseframework mit seinem Fokus und Teilsystemen präsentiert. Anschließend wird in Kapitel 3 die Datenbasis sowie die statistischen Voruntersuchung und Vorverarbeitung erläutert. Die Ergebnisse und werden anschließend in Kapitel 4 dargestellt und diskutiert. Die Arbeit schließt mit einem Fazit und Ausblick in Kapitel 5.

2. BESCHREIBUNG DES ANALYSEFRAMEWORKS

Datenbasierte Analyseframeworks für *Prognostics and Health Management* (PHM) Methoden können auf unterschiedliche Weise definiert bzw. gegliedert werden. SARKAR [5] beispielsweise unterteilt das so genannte *Health Monitoring* in drei Schritte: Die Fehlerdetektion, Diagnose und Prognose. KIM [4] beschreibt das PHM hingegen als die vier Schritte der Datenakquisition, Diagnose, Prognose und Gesundheitsmanagement. Eine ähnliche Strukturierung findet sich ebenso im Open System Architecture for Condition-Based Maintenance (OSA-CBM). Diese Arbeit wird sich Teilen dieser Ansätze bedienen und durch den Cross Industry Standard Process for Data Mining (CRISP-DM) ergänzen. Letzterer legt nahe, zunächst ein

ausführliches Datenverständnis durch Deskription zu erlangen.

Die wesentlichen Bestandteile des Analyseframeworks sind in BILD 2 dargestellt. Zu allererst wird die Datenakquisition durchgeführt. Relevante Daten müssen erfasst, kategorisiert und möglicherweise aus verschiedenen Quellen fusioniert werden. Auch die Deskription der Daten wird während der Akquisition durchgeführt. Im nächsten (optionalen) Schritt werden die Daten vorverarbeitet. Je nach Art der Daten können sie gefiltert, extrahiert oder reduziert werden. Dabei helfen Methoden wie die Principal Component Analysis (PCA) oder auch Methoden der Feature Selection. Nach der Vorverarbeitung folgt im Herzstück des Frameworks die Diagnose, bei der Fehler detektiert, isoliert, quantifiziert oder klassifiziert werden. Basierend auf diesen Informationen wird der aktuelle Gesundheitszustand des Systems beschrieben. Auf Grundlage der Diagnose erfolgt eine mögliche Prognose der Restlebensdauer (RUL) und schließlich eine Evaluation der Ergebnisse.

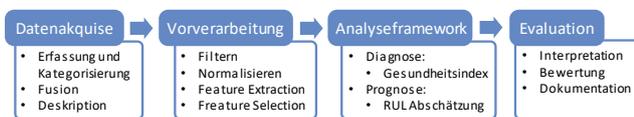


BILD 1. Schematische Darstellung des entwickelten Analyseframeworks

In diesem Beitrag wird der Fokus auf die Diagnose gelegt, wobei im Besonderen auf die Deskription der Daten und mögliche Einflüsse der Vorverarbeitung der Daten auf die Diagnose- und Prognoseergebnisse eingegangen wird. Zur Deskription der Daten werden Methoden der deskriptiven und explorativen Statistik angewandt. In der Datenvorverarbeitung sollen der Einfluss von Filtermethoden (gleitender Mittelwert) untersucht werden. Dieser Schritt wird im Rahmen von Kapitel 3 erläutert.

Die Diagnose des Systemzustands erfolgt in zwei Schritten: zunächst wird der Datensatz mithilfe einer Clusteranalyse in fünf verschiedene Klassen eingeteilt, die als diskreter Gesundheitsstatus (Health Index) verstanden werden können. Hierfür wird der K-Means Algorithmus aus der Familie der unüberwachten Lernverfahren (unsupervised learning) genutzt. Diese Vorsortierung ist notwendig um eine Datenbasis für die datenbasierte Klassifizierung zu schaffen. Im zweiten Schritt wird ein unkorrelierter Entscheidungswald Klassifizierer (Random Forest Classifier) auf die vorsortierten Daten angeleert. Dabei wird der aktuelle Systemzustand in vorhandene Klassen eingeordnet. Danach können nicht sortierte Testdaten anhand ihrer Parameterwerte in die verschiedenen Klassen eingeordnet werden und ihnen somit ein Gesundheitsstatus zugeordnet werden. Im letzten Schritt wird eine datenbasierte Regression angeleert, wieder basierend auf einem unkorrelierten Entscheidungswald Algorithmus, um die Prognose der Restlebensdauer zu ermöglichen.

Im Folgenden sollen die erwähnten Algorithmen kurz erklärt werden.

2.1 Verwendete Lernmodelle

K-Means Clustering

Im ersten Schritt werden die Daten durch einen unüberwachten Lernalgorithmus in Gruppen einsortiert und somit gruppiert (Clustering). Hierfür wurde ein K-Mittelwert Gruppieralgorithmus (im Folgenden K-Means Clustering oder K-Means Algorithmus) verwendet. Der K-Means Algorithmus, auch als Lloyds Algorithmus bekannt, wurde erstmals von LLOYD [9] beschrieben und gruppiert die Daten indem er verschiedene Stichproben in K Gruppen von gleicher Varianz verteilt und dabei ein Trägheitskriterium minimiert, auch bekannt als Summe der Quadrate innerhalb der Cluster. Die Anzahl an Gruppen K muss dabei zuvor spezifiziert sein.

Der K-Means Algorithmus teilt N Stichproben X in K getrennte Gruppen (Cluster) C , wobei jedes Cluster durch den Durchschnitt μ_j der Stichprobe im Cluster beschrieben wird. Die Durchschnitte werden im Allgemeinen als „Centroide“ bezeichnet. Der K-Means Algorithmus versucht dabei das Trägheits- (Inertia) oder auch das „Summe der Quadrate innerhalb der Cluster“ - Kriterium zu minimieren:

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_j - \mu_i\|^2) \quad (1)$$

Dabei arbeitet der Algorithmus im Wesentlichen in drei Schritten: Im ersten Schritt werden die Anfangscentroide ausgewählt. In diesem Beitrag wird hierfür die „k-means++“ Initialisierungsmethode nach ARTHUR [6] genutzt. Nach der Initialisierung durchläuft der Algorithmus Schleifen der Schritte zwei und drei. Zunächst wird jeder Datenpunkt dem nächstliegenden Centroid zugeordnet. Im dritten Schritt werden die neuen Centroide durch die Durchschnitte aller Datenpunkte im vorherigen Centroid berechnet. Die Differenz zwischen neuem und altem Centroid wird berechnet und die Schritte zwei und drei werden so lange wiederholt, bis die Differenz keine wesentlichen Änderungen mehr erfährt (siehe [8]).

Das Trägheitskriterium kann als Maß für die interne Kohärenz der Cluster verstanden werden, wobei folgende Vereinfachungen getroffen werden: Die Trägheit macht die Annahme, dass Cluster konvex und isotropisch sind, was nicht immer der Fall ist. Vor allem längliche Cluster oder Gruppierungen mit ungleichmäßigen Formen werden schlecht abgebildet (siehe [7]). Weiterhin ist die Trägheit keine normalisierte Metrik. Kleine Werte sind besser, aber in hochdimensionalen Räumen tendieren die euklidischen Distanzen zu überhöhen. Dimensionsreduktionsmethoden, wie die Hauptkomponentenanalyse können dabei entgegenwirken und die Performanz verbessern.

Random Forest Klassifizierung

Der unkorrelierte Entscheidungswald Klassifizierer (Random Forest Classifier) wurde zuerst von BREIMANN [5] beschrieben. Der Algorithmus basiert dabei auf *bagging* oder *bootstrap aggregation* in Kombination mit Entscheidungsbäumen. Bei Regressionsproblemen wird lediglich derselbe Entscheidungsbaum auf *Bootstraps* der Trainingsdaten angewandt und hinterher ein Durchschnitt des Ergebnisses genommen, bei Klassifikationsproblemen gibt jeder Baum eines *Komitees* eine Stimme für die prädierte Klasse ab und das Ergebnis wird in einem

Mehrheitsentscheidungsverfahren getroffen (siehe [8]).

Bei Random Forests wird ähnlich zu *bagging* eine große Sammlung an unkorrelierten Entscheidungsbäumen genutzt und das Ergebnis hinterher gemittelt. In vielen Fällen ist die Performanz vergleichbar zu *boosting*, jedoch ist der Random Forest einfacher abzustimmen. Da der Algorithmus auf Entscheidungsbäumen basiert, sollen diese im Folgenden kurz erläutert werden.

Entscheidungsbaumverfahren stellen einfache aber dennoch mächtige Werkzeuge für Klassifikations- oder Regressionsprobleme dar. Die Ergebnisse lassen sich gut visualisieren und nachvollziehen. An jedem Knoten (auch Attribut genannt) wählt das Lernverfahren ein Merkmal anhand der Beschreibungsvariablen der Trainingsdaten aus, welches den Datensatz am besten teilt. An jedem internen Knoten werden Entscheidungsregeln gelernt. Jeder Knoten stellt eine binäre Klassifikation dar, an dem zwei Nachfahren entstehen. Ein Ende eines Astes wird als Punkt ohne weitere Teilung definiert und als Blatt bezeichnet. An diesen Blättern werden die Regressionsergebnisse als Klassen ausgegeben (siehe [10]).

Die Güte der Baumarchitektur kann als Maß des Informationsgehalts beschrieben werden. Hier kann die Entropie E mit

$$E(P) = E(P_1, P_2, \dots, P_n) = \sum_{i=1}^n P_i \cdot \log_2(P_i) \quad (2)$$

als Maß für den Informationsgehalt genutzt werden. Diese stellt auch eine Wahrscheinlichkeitsverteilung für die verschiedenen Ausgänge des Baumes dar. Gleichzeitig stellt die Entropie ein Maß für die Unsicherheit dar. Der Informationsgehalt I einer Datenmenge S ist definiert als:

$$I(S) = 1 - E(S). \quad (3)$$

Der Informationsgewinn G durch die Verwendung eines Merkmals A an jedem Knoten wird über

$$G(S, A) = E(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot E(S_i) \quad (4)$$

berechnet. Dabei stellt der zweite Teil von Gleichung (4) den Informationsgehalt durch das Merkmal A in den Teilmenge S_i dar. Der Informationsgehalt wird bei der Teilung des Baumes an den Knoten verwendet. Ein Merkmal ist umso besser, je mehr seine Aufteilung den Informationsgehalt der Verteilung erhöht (siehe [10] und [11]).

Der Aufbau und Ablauf von Random Forest Algorithmen läuft in der folgenden Schleife der Länge B, bestehend aus zwei Schritten, ab: Zunächst wird eine „Bootstrap-Stichprobe Z^* “ der Größe N aus dem Trainingsdatensatz entnommen. Im zweiten Schritt wie ein Zufallsbaum T_b zu Z^* nach folgendem Schema aufgebaut bis eine Minimalknotengröße erreicht ist. Für jeden Endknoten werden zufällig aus p Variablen m ausgewählt, der beste Teilungspunkt wird identifiziert und der Knoten in zwei weitere Nachfahren geteilt. Danach werden die einzelnen

Bäume zu einem Gesamtsystem $\{T_b\}_1^B$ der Bäume zusammengeführt. Im letzten Schritt wird für eine Regression ein neuer Wert x wie folgt vorhergesagt:

$$\hat{f}_{\text{RF}}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (5)$$

Für die Klassifikation wird $\hat{C}_b(x)$ als Prädiktion der Klasse des b-ten Entscheidungsbaumes die Gesamtklassifikation von x durch $\hat{C}_{\text{RF}}^B(x) = \text{Mehrheitsentscheidung}\{\hat{C}_b(x)\}_1^B$ beschrieben (siehe [8]).

Ausgehend von der Beschreibung des Analyseframeworks und der verwendeten Lernalgorithmen wird nun auf die Datenbasis eingegangen.

3. DATENBASIS

Die Analyse dieses Beitrags basiert auf einem Datensatz der vom National Aeronautics and Space Administration (NASA) Prognostics Center of Excellence (PCoE) bereits im Jahr 2008 zur Verfügung gestellt wurde und noch heute zum Download zur Verfügung steht (siehe [1]). Der verwendete Datensatz wird als „Turbofan Degradation Simulation Data Set“ bezeichnet und umfasst insgesamt 12 Datensätze die wie folgt kategorisiert werden können: Zunächst teilt sich der Datensatz in Trainings- und Testdaten sowie den Testdatensätzen zugehörigen Restlebensdauerwerten. Während im Trainingsdatensatz vollständige *run-to-failure* Daten vorliegen, mit denen Prognosealgorithmen entwickelt werden können, beinhaltet der Testdatensatz keine Daten bis zum vollständigen Versagen, jedoch wird die tatsächliche Restlebensdauer als Information mitgegeben.

Weiterhin wird der Datensatz in vier unterschiedliche Teildatensätze unterteilt, die sich in ihrer Komplexität hinsichtlich eines Prognosealgorithmus unterscheiden. Je nach Datensatz liegen einer oder mehrere Betriebszustände oder Versagensfälle vor.

Eine Übersicht der verschiedenen Datensätze ist in TAB 1 dargestellt.

Datensatz	Betriebszustände	Versagensfälle	Trainings-einheiten	Test-einheiten
1	1	1	100	100
2	6	1	260	259
3	1	2	100	100
4	6	2	249	248

TAB 1. Übersicht des Datensatzes

In diesem Beitrag wird zunächst nur auf Datensatz 1 eingegangen. Während der statistischen Voruntersuchung ergab sich, dass die anderen, komplexen Datensätze lediglich aufwendiger vorverarbeitet werden müssen, bspw. durch die Bereinigung der Betriebszustands, dann aber gleiche Charakteristika wie Datensatz 1 aufweisen.

Jeder Datensatz enthält dabei eine unterschiedliche Anzahl an Trainings- und Testeinheiten. Jede Trainingseinheit kann als Zeitreihe eines einzelnen Triebwerks verstanden werden, womit alle Einheiten als eine Flotte von Triebwerken gleichen Typs interpretiert werden können. Jedes Triebwerk weist dabei unterschiedliche anfängliche Abnutzung und Herstellerabweichungen auf, die dem

Anwender nicht bekannt sind und normalverteilt sind. Das Triebwerk wird normal betrieben und entwickelt einen Fehler an einem Punkt in der Zeitreihe.

Neben den drei Variablen zur Beschreibung des Betriebszustandes werden 21 Sensorparameter aufgezeichnet. Die Zeiteinheit wird als Zyklus beschrieben und kann im erweiterten Sinne als eine Messung pro Flug verstanden werden.

Weiterhin würden bei realen Triebwerken Prozess- und Messrauschen die Ausgaben der Sensoren überlagern. In der Simulation werden die Signale mit verschiedenen Rauschwerten überlagert um dieses Rauschen darzustellen. Exemplarisch wird der Verlauf des Sensors 4 (Temperatur am Niederdruckturbinenaustritt) über die Lebensdauer von Triebwerk Nr. 3 in BILD 2 dargestellt. Während der hier gezeigt Verlauf schon einen charakteristischen Verlauf entsprechend von Degradationsverläufen aufzeigt, bleiben andere Sensoren weitgehend konstant. Aus diesem Grund wird im Folgenden eine ausführliche Voruntersuchen der Daten vorgenommen.

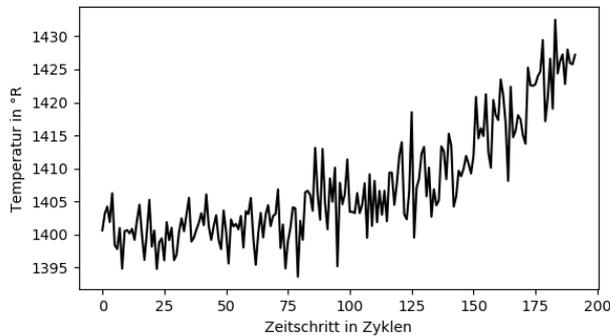


BILD 2. Exemplarischer Verlauf von Sensor 4 (Temperatur am Niederdruckturbinenaustritt) über die Lebensdauer eines Triebwerks

3.1 Statistische Voruntersuchungen

Zur systematischen Untersuchung des Datensatzes und zum fundierten Verständnis der Datengrundlage wird eine statistische Voruntersuchung der Daten vorgenommen. Dabei werden vor allem Methoden der Deskriptiven und Explorativen Statistik angewandt. Mithilfe der Beschreibung der Daten durch Kennzahlen und Darstellungen in charakteristischen Grafiken wird das Datenverständnis erhöht. So stehen diese Methoden häufig vor dem nächsten Schritt der induktiven Statistik (hier umgesetzt durch Methoden der künstlichen Intelligenz) und erlauben im ersten Schritt eine Beschreibung des Datensatzes und das inhaltliche Kennenlernen (siehe [3]).

In diesem Kapitel werden beispielhaft einige der Ergebnisse der deskriptiven und explorativen Voruntersuchung präsentiert um ausgehend davon bestimmte Parameter für intelligente Lernmodelle ausschließen zu können. Wesentliche Kenngrößen sind hier die Streumaße einzelner Parameter sowie Korrelationen zwischen verschiedenen Parametern.

Im ersten Schritt der deskriptiven Statistik werden klassische beschreibende Merkmale, wie Lage- und Streumaße der Parameter, untersucht. Beispielhaft werden in BILD 3 Histogramme für zwei der aufgezeichneten

Sensorwerte gezeigt. So zeigt die physikalische Fan-Geschwindigkeit (Sensor 8) eine schiefe Normalverteilung während der absolute Druck im Bypasskanal (Sensor 6) lediglich zwei Ausprägungen aufweist, die weiterhin annähernd gleich sind (Standardabweichung = 0.0014). Mithilfe dieser Information lässt sich bspw. Sensor 6 für die Verwendung im Diagnose- bzw. Prognoseframework ausschließen.

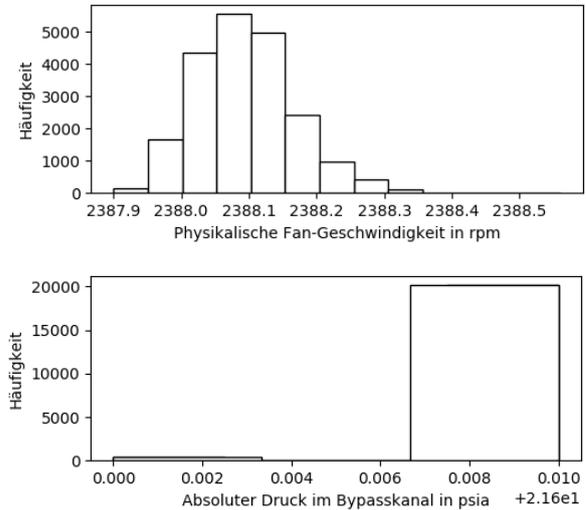


BILD 3. Beispielhafte Ausprägung von Sensor 8 (physikalische Fan-Geschwindigkeit) und Sensor 6 (absoluter Druck im Bypasskanal)

Ausgehend hiervon können innerhalb der explorativen Analyse Korrelationen zwischen einzelnen Parametern untereinander aber auch im Hinblick auf die RUL untersucht werden. Streudiagramme ermöglichen eine Visualisierung dieser Effekte. In BILD 4 werden die Streudiagramme zwischen physikalischer Kerngeschwindigkeit (Sensor 9) und Restlebensdauer bzw. korrigierte Kerngeschwindigkeit (Sensor 14) dargestellt. Beide Streudiagramme zeigen unterschiedliche Charakteristiken, die jeweils einen Informationsgewinn bedeuten. So zeigt sich zwischen Sensor 9 und Restlebensdauer (RUL) beispielweise keine eindeutige Korrelation, da der Verlauf mit abnehmender Restlebensdauer entweder zunehmen oder abnehmen kann. Zwischen Sensor 9 und Sensor 14 scheint ein eindeutiger, linearer Zusammenhang zu bestehen, was die hohe Korrelation (nach Pearson) von 0,963 zusätzlich bestätigt.

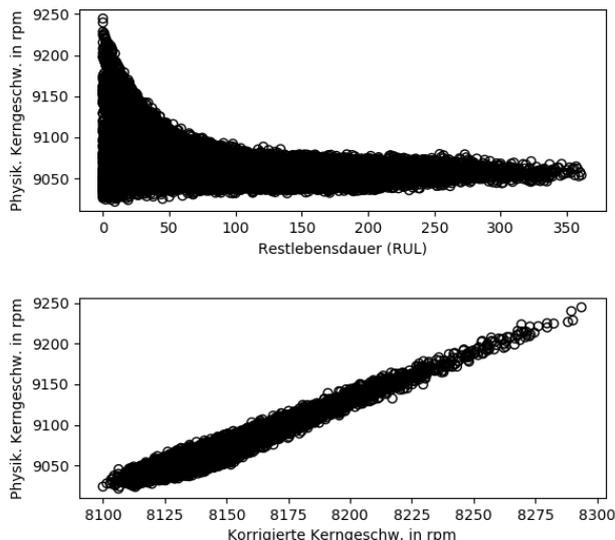


BILD 4. Streudiagramme zwischen Kerngeschwindigkeiten und Restlebensdauer

Neben diesen Beispielen finden sich sechs weitere Sensorpaare, die eine starke Korrelation ($|r| > 0,8$) aufweisen. Konstante Parameter weisen sehr geringe Korrelationen mit der RUL auf und beinhalten bezüglich der Degradation des Triebwerks keine Information. Weiterhin besteht nach der Korrelationsanalyse kein Zusammenhang zwischen allen Betriebsparametern und Sensorparametern bzw. Restlebensdauer. Dies wird vor allem durch die weitgehend konstanten Betriebsparameter begründet.

Bezüglich der physikalischen Bedeutung, bestätigt die Korrelationsanalyse die Angaben der Parameterliste von SAXENA [2]. Basierend auf den Ergebnissen der statistischen Voruntersuchung werden sieben unabhängige Parameter ausgewählt um im Analyseframework als Eingangsparameter zu dienen. Diese sind in TAB 2 zusammengefasst.

Sensor-Nr.	Aufgezeichneter Parameter
2	Temperatur am Niederdruckverdichteraustritt
3	Temperatur am Hochdruckverdichteraustritt
4	Temperatur am Niederdruckturbinenaustritt
7	Totaldruck am Hochdruckverichteraustritt
11	Statischer Druck am Hochdruckverichteraustritt
12	Treibstoffflussverhältnis
15	Bypass-Verhältnis

TAB 2. Für das Analyseframework verwendete Parameter

Dabei stellen diese Parameter die Grundlage für die Clusteranalyse und Klassifikation dar. Für die Prognose der Restlebensdauer innerhalb der Regression wird weiterhin noch die Zyklenanzahl (Zeitschritt) und der Gesundheitsindex verwendet.

3.2 Datenvorverarbeitung

Als optionaler Bestandteil des Analyse Frameworks besteht die Möglichkeit eine Datenvorverarbeitung vorzunehmen. Da die vorliegenden Sensorparameter künstlich stark verrauscht wurden (vgl. BILD 2), bietet sich eine Vorverarbeitung der Daten an. Die Daten weisen dabei

keine Ausreißer auf, sondern unterliegen lediglich hochfrequenten, künstlichen Rauschfunktionen.

Der erste Schritt der Datenvorverarbeitung sieht somit eine Glättung vor. Hierfür werden im Allgemeinen Tiefpassfilter verwendet und es stehen eine Vielzahl verschiedener Filtermethoden zur Verfügung. Im Rahmen dieses Beitrags soll nur der Einfluss einer Methode exemplarisch dargestellt werden. Hierfür wurde der gleitende Mittelwert (Moving Average) verwendet. Der gleitende Mittelwert bildet den Mittelwert über eine Reihe von Eingangssignalen s_{i+j} um ein Ausgangssignal y_i zu berechnen. Diese Prozedur wird wiederholt, wobei immer der Mittelwert über ein gleitendes Fenster mit M Messpunkten berechnet wird. Die folgende Gleichung beschreibt den mathematischen Zusammenhang:

$$y_i = \frac{1}{M} \sum_{j=0}^M s_{i+j} \tag{6}$$

Mit der Glättung der Daten geht jedoch auch ein Informationsverlust einher. So wird durch den gleitenden Mittelwert ein Teil der Datenmenge abgeschnitten. BILD 5 zeigt beispielhaft den Sensorverlauf der Temperatur am Niederdruckturbinenaustritt über einen Lebenszyklus (vgl. BILD 2) mit ungefilterten Originaldaten und gefiltert durch die gleitende Mittelwert-Methode.

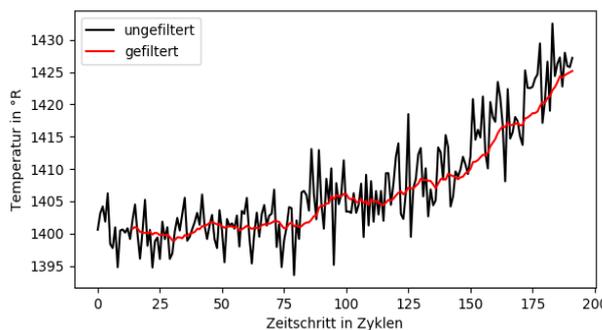


BILD 5. Beispielhafte Darstellung des Einflusses von Filtermethoden auf die Temperatur am Niederdruckturbinenaustritt

Der allgemeine Trend in den Daten wird gut abgebildet, hochfrequente Anteile und Rauschteile werden gefiltert, wobei jedoch die Dynamik des Systems noch hinreichend gut abgebildet wird. Die iterative Auswahl der Fensterbreite von 15 Zyklen hat für diese Filtermethode dabei die besten Ergebnisse im Hinblick auf die Klassifikation erreicht.

3.3 Trainings- und Testdaten

Bei datenbasierten Lernverfahren wird zwischen Trainings- und Testphase unterschieden. In der Trainingsphase werden die Knoten und Kanten des Entscheidungsbaumes definiert. Dabei kennt der Algorithmus den wahren Wert der Zielgröße und kann dahingehend seine Parameter optimieren. In der Testphase wird anhand des Testdatensatzes geprüft, wie gut der Algorithmus die Zielgröße abbilden kann. Die Zielgröße ist hierfür bekannt, wird jedoch zurückgehalten und dann mit dem vom Algorithmus berechneten Werten verglichen. Zu Beginn der Berechnung wird somit der gesamte zur Verfügung stehende Datensatz aufgeteilt in Test- und Trainingsdaten.

In diesem Beitrag wurden 80% des Datensatzes zum Training verwendet und 20% zum Testen. Dabei wurde der Datensatz zufällig geteilt und gemischt (randomisiert). Ein zweiter Testdatensatz wird durch das NASA PCoE bereit gestellt, unterscheidet sich jedoch von Testdatensatz, der aus den Trainingsdaten entnommen wurde: Der Testdatensatz der NASA beinhaltet 100 Schadensverläufe von Triebwerken, die zu einem beliebigen, jedoch meist recht frühen, Zeitpunkt abgeschnitten sind. Damit weisen die meisten Messpunkte noch keine fortgeschrittene Degradierung auf. Die aus den Trainingsdaten entnommenen Testdaten sind hingegen zufällig ausgewählt und enthalten somit Messpunkte aus dem gesamten Lebenszyklus. Der Testdatensatz, welcher aus dem Gesamtdatensatz entnommen wurde, wird im Folgenden als Testsatz 1 bezeichnet, der zusätzliche Testdatensatz der NASA als Testsatz 2.

3.4 Validierungsmetriken

Zur Validierung und Bewertung der Ergebnisse der Prognose der Restlebensdauer werden nachfolgende Metriken benutzt. Der Mittlere Absolute Fehler (MAF) repräsentiert die absolute Abweichung zwischen prädiiziertem Ergebnis und realem Modellierungstarget gemittelt über die Anzahl der Beobachtungen. Der Mittlere Relative Fehler (MRF) ist der auf den Mittelwert der Stichprobe bezogene mittlere absolute Fehler. Der Mittlere Quadratische Prognosefehler (MQF) ist die Quadratwurzel des mittleren, quadratischen Fehlers. Der Determinationskoeffizient (R^2) ist ein Gütemaß zur Erklärung der Datenvariation im Regressionsmodell. Die verwendeten Metriken werden in SAXENA [12] detailliert beschrieben.

4 ERGEBNISSE UND DISKUSSION

Im Folgenden werden zunächst die Ergebnisse aus der Clusteranalyse und der Klassifikation der Daten präsentiert, gefolgt von der Darstellung der Ergebnisse der Regression. Abschließend werden die Ergebnisse diskutiert.

BILD 6 zeigt die Einordnung der Clusteranalyse beispielhaft an Triebwerk Nr. 3 in fünf Klassen. Die Klassenzahl wurde dabei normiert, sodass eine Übertragung auf einen Gesundheitsstatus möglich ist. Folglich sind Triebwerke bei Gesundheitsstatus 0 gesund und degradieren zunehmend bis sie bei Gesundheitsstatus 1 nahezu kaputt sind.

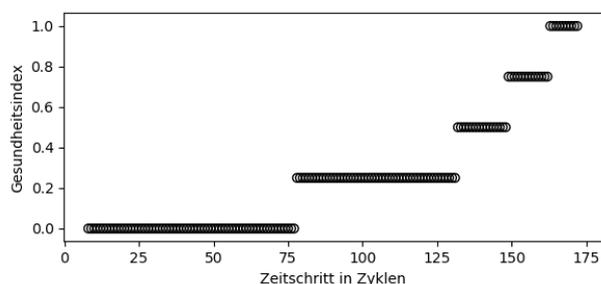


BILD 6. Beispielhafte Darstellung des Verlaufs des Gesundheitsindex anhand von Triebwerk Nummer 3

Der ansteigende Trend des Gesundheitsindex ähnelt dabei einer Exponentialfunktion, die häufig für Degradationsprozesse genutzt wird. In der Clusteranalyse wird die komplette Datenbasis verarbeitet, was auch dazu führen kann, dass einige Triebwerke nicht alle Klassen

aufweisen. Der Gesundheitsindex muss somit nicht immer zu Beginn in Klasse 0 einsortiert werden. Weiterhin kann sich der Gesundheitsindex auch abwärtspringen, was als temporäre Verbesserung des Gesundheitsstatus interpretiert werden kann. BILD 7 zeigt exemplarisch das Ergebnis der Clusteranalyse für die Triebwerke 38 und 68, die eine solche Charakteristik aufweisen.

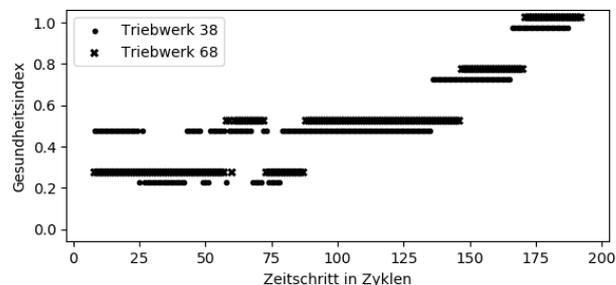


BILD 7. Beispielhafte Darstellung des Verlaufs des Gesundheitsindex von Triebwerken Nr. 38 und 68

Im nächsten Schritt wird der unkorrelierte Entscheidungswald Klassifizierer mit den nun gelabelten Daten angeleert. Basierend auf den Testdaten lässt sich die Performanz des Algorithmus' beschreiben. Diese lässt sich anhand der Korrektklassifikationsrate (Accuracy) und Sensitivität (Recall) beschreiben. Weiterhin lässt sich mit der Wahrheitsmatrix (Confusion Matrix) beschreiben, wie das Verhältnis von prädiizierten Klassen zu tatsächlichen Klassen ist. Diese ist in BILD 8 normalisiert dargestellt.

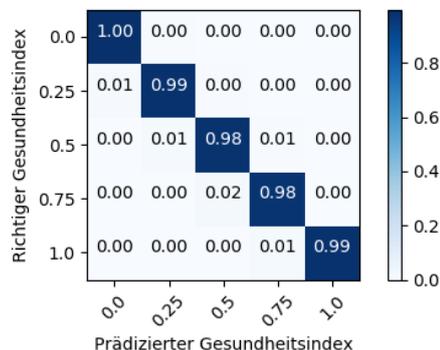


BILD 8. Normalisierte Wahrheitsmatrix zur Ergebnisdarstellung des Klassifizierers

Es lässt sich erkennen, dass die Gesundheitszustände über den gesamten Lebenszyklus gleichmäßig auf sehr hohem Niveau klassifiziert werden. Dieses Ergebnis wird auch durch die Korrektklassifikationsrate von 98,5% und die Sensitivität von 99% widerspiegelt. Die Diagnosegenauigkeit ist damit sehr hoch, der aktuelle Systemzustand kann basierend auf der Logik der Clusteranalyse gut klassifiziert werden. Dies bildet eine wichtige Basis für die Prognose, die im Allgemeinen mit größeren Unsicherheiten behaftet ist.

Im letzten Schritt lassen sich die Ergebnisse der Regression darstellen. Dabei wird zunächst auf die Ergebnisse des Testdatensatz 1 eingegangen. BILD 9 zeigt hierfür die Ergebnisse der Prädiktion, wobei die wahre Restlebensdauer (Target) auf der Abszisse aufgetragen wird und die durch den Algorithmus prädiizierte Restlebensdauer (Prediction) auf der Ordinate abgebildet ist. Die Winkelhalbierende markiert die perfekte Prädiktion,

in der Prädiktion und wahrer Wert theoretisch identisch wären.

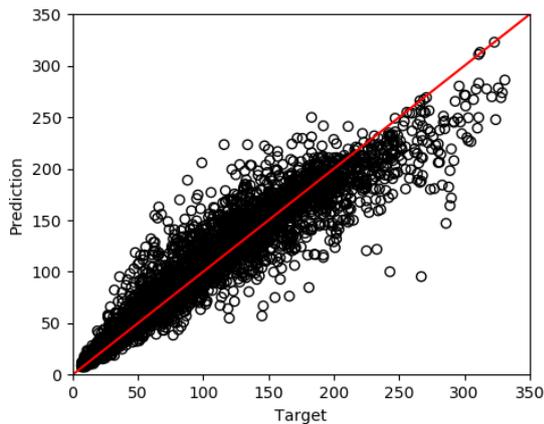


BILD 9. Reale und mittels unkorreliertem Entscheidungswald prädizierte Restlebensdauer

Entscheidend für bei der Prädiktion der Restlebensdauer ist der konvergierende Verlauf, der sich in dieser Darstellung wiederfinden lässt. Der Fehler bei der Prädiktion wird also mit abnehmender Restlebensdauer kleiner bzw. die Prognose genauer. Weiterhin ist erkennbar, dass der Algorithmus bei Restlebensdauern größer als etwa 200 Zyklen zu niedrig schätzt. Eine Unterschätzung der Restlebensdauer wird im Allgemeinen einer Überschätzung vorgezogen, weshalb diese Abweichung nicht zu kritisch bewertet werden muss. Jedoch wird Gesamtergebnis negativ beeinflusst.

BILD 10 zeigt die Ergebnisse der Prädiktion von Testdatensatz 2, der deutlich mehr Datenpunkte aufweist. Auffällig ist, dass hier keine Datenpunkte nahe Restlebensdauer gleich null (kaputt) liegen, da die Datenbasis wie zuvor beschrieben bereits vorzeitig abgeschnitten wird.

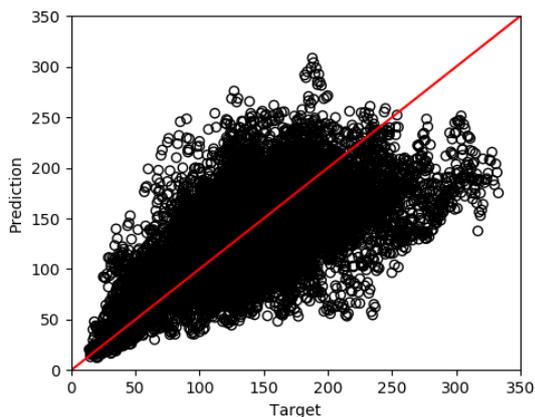


BILD 10. Reale und mittels unkorreliertem Entscheidungswald prädizierte Restlebensdauer anhand von Testdatensatz 2

Ähnlich wie bei Testdatensatz 1 lässt sich ein konvergierender Trend in der Prognose erkennen, jedoch streuen die Prognosewerte deutlich mehr. Auch die Validierungsmetriken der Prognose weisen schlechtere Werte auf. Diese Werte sind in TAB 3 zusammengefasst.

	Testdatensatz 1	Testdatensatz 2
MAF [Zyklen]	14,7	34,1
MRF [%]	7,2	16,6
MQF [Zyklen]	22,3	44,6
R ² [-]	0,884	0,382

TAB 3. Vergleich der Validierungsmetriken der Prognose

Die Ergebnisse des Testdatensatz 1 sind durchaus vergleichbar mit den Ergebnissen diverser Veröffentlichungen im Rahmen der NASA PHM Data Challenge 2008. Einige der Veröffentlichungen wurden von RAMMASSO [13] zusammengefasst. Hierfür wurde ein anderer, jedoch vergleichbarer Datensatz verwendet, der dem hier analysierten Datensatz um die verschiedenen Betriebszustände erweitert wurde (vgl. Datensatz 2 aus TAB 1). Der mittlere relative Fehler, sowie der Determinationskoeffizient weisen auf hinreichend gute Genauigkeiten hin. Eine mögliche Begründung für die schlechteren Validierungsmetriken in Testdatensatz 2 ist das Vorhandensein von Datenpunkten von überwiegend gesunden Triebwerken. Die Verteilung des Gesundheitsindex über den gesamten Testdatensatz 2 zeigt, dass die meisten Datenpunkte einen Gesundheitsindex zwischen 0 und 0,5 aufweisen, während bei Testdatensatz 1 gleichmäßig viele Datenpunkte über den gesamten Degradationsprozess vorhanden sind. Weiterhin hat sich gezeigt, dass das Mischen der Datenbasis (Shuffling) für das Training des Regressionsmodells einen maßgeblichen Einfluss hat. Wird dies nicht vorgenommen, erreicht das Framework im besten Fall vergleichbare Validierungsmetriken wie bei Testdatensatz 2. Der maßgebliche Einfluss hiervon ist jedoch nicht abschließend geklärt.

Bei weiteren Untersuchungen mit geringerer Anzahl an Eingangsparametern in das Lernmodell verschlechtert sich die Genauigkeit wie auch der Determinationskoeffizient von Testdatensatz 1 erwartungsgemäß, die Validierungsmetriken für Testdatensatz 2 verbessern sich jedoch. Auch dieses Phänomen soll in weiteren Arbeiten genauer untersucht werden.

Die Ergebnisse der Diagnose des aktuellen Systemzustands, welche in dieser Arbeit im Vordergrund stehen sollten, weisen durchweg sehr hohe Sensitivität und Korrektklassifikationsraten auf. Besonders im Abschnitt voranschreitender Degradation (Gesundheitsindex > 0,5) sind keine Sprünge zwischen den Klassen mehr zu erwarten. Bei überwiegend gesunden Triebwerken mit höheren Restlebensdauern sind hingegen einzelne Sprünge in den Gesundheitszuständen identifiziert worden. Gerade im Hinblick auf die eingangs erwähnte Tatsache, dass alle Triebwerke zu Beginn der Messreihe zu 99% gesund sind, ist dieses Verhalten nicht repräsentativ und soll noch weiter untersucht werden. Hierbei kann eine Anpassung der Parameter der Clusteranalyse möglicherweise Verbesserung schaffen.

Weiterhin können auch verschiedene Methoden der Datenvorverarbeitung, wie die Hauptkomponentenanalyse bessere Ergebnisse hervorrufen. Erste Ergebnisse zeigen, dass schon zwei Hauptkomponenten ausreichen, um 95% der Varianz der Parameter zu erklären. Die Auswirkungen auf das Analyseframework sind jedoch noch nicht detailliert untersucht worden.

Ebenfalls wäre es interessant den zeitvarianten Einfluss der

Parameterverläufe auf die Restlebensdauer oder auch den Gesundheitsindex zu untersuchen. Hierfür wird gerade eine mögliche Implementierung von rekurrenten Neuronalen Netzen untersucht.

5 FAZIT UND AUSBLICK

Im vorgestellten Beitrag wird eine Analyseframework zu Analyse von Triebwerkdegradationsdaten mithilfe von maschinellen Lernalgorithmen vorgestellt. Neben einer statistischen Voruntersuchung der Datenbasis mittels deskriptiven und explorativen Methoden der Statistik zu Identifizierung relevanter Parameter, besteht das Herzstück des Frameworks aus eine Diagnose- und Prognosesystem. Dafür wurde eine Clusteranalyse, basierend auf einem K-Means Algorithmus genutzt um die Daten zunächst zu gruppieren. Mit einem unkorreliertem Entscheidungswald Klassifizierer wird dann der aktuelle Systemzustand diagnostiziert und davon ausgehend mithilfe eines unkorreliertem Entscheidungswald Regressors die Restlebensdauer prognostiziert. Die Diagnose liefert dabei gute Genauigkeiten bei der Klassifikation. Die Prognose der Restlebensdauer soll noch weiter untersucht werden.

Im Hinblick auf die Diagnose des Systemzustandes im Zusammenspiel mit der Prognose der Restlebensdauer ließe sich auch die Potenziale eines binären Klassifikationsansatzes untersuchen. Beispielsweise könnte man klassifizieren, ob ein System zu einem aktuellen Zustand noch mehr als 30 Zyklen Restlebensdauer bis zum Versagen aufweist, oder weniger. Diese Information wäre für einige Anwendungen schon hinreichend genau, und mit höherer Zuverlässigkeit berechenbar.

Auch eine Ausweitung der Analyse auf die in TAB 1 genannten Datensätze zur Vergleichbarkeit mit den Ergebnissen der NASA PHM Data Challenge von 2008 wird angestrebt.

Weiterhin wird an einer Verwendbarkeit dieses Beitrages für das Forschungsvorhaben *INDI* (Intelligente Datennutzung in der Instandhaltung) im aktuellen Luftfahrtforschungsprogramm des Bundes gearbeitet. Im Rahmen von *INDI* sollen Wartungsdaten aus verschiedenen Datenbanken des Forschungspartners Lufthansa Technik AG genutzt werden, um in einem Datenanalyse-Framework sinnvoll gefiltert, miteinander verknüpft und intelligente analysiert werden. Somit soll die Flugzeugwartung durch neue, digitale und datenbasierte Produkte unterstützt werden. Im Hinblick auf *INDI* bildeten die zugrunde liegenden Daten eine mögliche Datenbasis und lassen sich im entfernten mit ACMS-Reports (Aircraft Condition Monitoring System) vergleichen. Diese übertragen bei zuvor bestimmten Auslösebedingungen einen Schnappschuss der aktuellen Messwerte ausgewählter Komponenten im Flugzeug an den Wartungsbetrieb. Die Datenbasis soll durch weitere Datenquellen des Wartungsbetriebs ergänzt werden. Dabei stehen vor Allem Datenfusionsstrategien im Vordergrund.

DANKSAGUNG

Die Autoren bedanken sich an dieser Stelle beim Bundesministerium für Wirtschaft und Energie (BMWi) für die Unterstützung im Teilvorhaben *INDI* (Intelligente Datennutzung in der Instandhaltung) zur Entwicklung intelligenter Algorithmen und Datenanalyse-Strategien aus

dem Bereich des maschinellen Lernens zur Optimierung in der Flugzeuginstandhaltung im Rahmen des Luftfahrtforschungsprogramms (LuFo V-3) an der TU Darmstadt. Neben dem Institut für Flugsysteme und Regelungstechnik findet sich im Projekt *INDI* der Projektpartner Lufthansa Technik AG. Das Projekt endet am 30.06.2021.

KONTAKTADRESSEN

Henrik Simon, M. Sc.
simon@fsr.tu-darmstadt.de

Prof. Dr.-Ing. Uwe Klingauf
klingauf@fsr.tu-darmstadt.de

LITERATURANGABEN

- [1] National Aeronautics and Space Administration (Hrsg.): NASA Prognostics Data Repository. Abrufbar unter:
<https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/> (abgerufen am 31.05.2018)
- [2] Saxena, A., Goebel, K., Simon, D. und Eklund, N.: „Damage propagation modeling for aircraft engine run-to-failure simulation“ (2008)
- [3] Becker, T., Herrmann, R., Sandor, V., Schäfer, D. und Wellisch, U.: „Deskriptive Statistik und explorative Datenanalyse“ in Stochastische Risikomodellierung und statistische Methoden, Berlin, Heidelberg: Springer Berlin Heidelberg S. 27–91, (2016)
- [4] Kim, N.-H., An, D. und Choi, J.-H.: „Prognostics and Health Management of Engineering Systems“, Springer International Publishing, Schweiz
- [5] Breiman, L., „Random forests“ in Machine learning, 45, S. 5-32, (2001)
- [6] Arthur, D. and Vassilvitskii, S. „K-Means++: The Advantages of Careful Seeding“ in: Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms, S. 1027-1035 (2007)
- [7] Aggarwal, C.C. „Data mining: the textbook“ Springer (2015)
- [8] Friedman, J., Hastie, T. and Tibshirani, R. „The elements of statistical learning“ Vol. 1. No. 10. New York, USA: Springer series in statistics (2001)
- [9] Lloyd, S. „Least squares quantization in PCM“ IEEE transactions on information theory, S. 129-137 (1982)
- [10] Moisen, G.G.: „Classification and Regression Trees“ In: Encyclopedia of Ecology, Vol. 1, S. 582-588 (2008)
- [11] Ertel, W.: „Grundkurs Künstliche Intelligenz“ Springer Vieweg Wiesbaden (2016)
- [12] Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S. und Schabacher, M.: „Metrics for Evaluating Performance of Prognostics Techniques“, Institute of Electrical and Electronics Engineers IEEE, International Conference on Prognostics and Health Management PHM (2008)
- [13] Ramasso, E., Saxena, A.: „Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets“ In: International Journal of Prognostics and Health Management (2014)