

Abschlussbericht

30.11.2020

Entwicklung einer neuen Methode zur Ressourceneinsparung durch prozessorientiertes, digitales Stoffstrommanagement in produzierenden mittelständischen Unternehmen (Nova) – FKZ: 34589/10

Hochschule für Technik und Wirtschaft Berlin

Prof. Dr. Volker Wohlgemuth
Studiengang Betriebliche Umweltinformatik
Wilhelminenhofstraße 75A
12459 Berlin



NOVAPAX Kunststofftechnik Steiner GmbH & Co. KG
Umweltmanagement, Arbeitssicherheit
Schätzelbergstraße 4-10
12099 Berlin

NOVAPAX

Inhaltsverzeichnis

1	Problemstellungen und Hauptzielstellung	4
2	Wissenschaftlich-technische Ergebnisse	5
2.1	Datenvorverarbeitung und -analyse.....	5
2.2	Modellgenerierung und -analyse	11
2.3	Ergebnisse der technischen Umsetzung und Handlungsempfehlungen	21
2.3.1	Interpretation der Diagramme	21
2.3.2	Handlungsempfehlungen Maschine 64	22
2.3.3	Handlungsempfehlungen Maschine 67	24
2.3.4	Handlungsempfehlungen Maschine 68	25
2.3.5	Handlungsempfehlungen Maschine 69	26
2.3.6	Allgemeine Handlungsempfehlungen.....	27
2.3.7	Entwicklung eines künstlichen neuronalen Netzes (KNN)	28
3	Mögliche Anschlussarbeiten	34
4	Fortschreibung des Verwertungsplanes.....	35
5	Während des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt	36
6	Anhänge	37

Abbildungsverzeichnis

Abbildung 1 Training des KNN.....	29
Abbildung 2 Modell der Qualitätsverluste der Maschine M64.....	30
Abbildung 3 Trainings- und Validationsverlust.....	31
Abbildung 4 Trainings- und Validierungsverlust - Varianzreduktion.....	32
Abbildung 5 Modellgenauigkeiten des Qualitäts- und Energiemodells	33

Tabellenverzeichnis

Tabelle 1 Ergebnisse der Klassifizierung für Maschine M64.....	30
Tabelle 2 M64: Qualitäts- und Energieempfehlung und prognostizierte Verbesserung.....	33
Tabelle 3 M67: Qualitäts- und Energieempfehlung und prognostizierte Verbesserungen.....	34
Tabelle 4 M69: Qualitäts- und Energieempfehlung und prognostizierte Verbesserungen.....	34

Zusammenfassung

Im von der HTW in Kooperation mit dem Kunststoffhersteller Novapax Kunststofftechnik Steiner GmbH & Co. KG Projekt Nova wurde untersucht, ob und inwieweit Machine Learning zur Optimierung des Energieverbrauchs sowie der Reduktion von fehlproduzierten Kunststoffteilen in einem kleinen und mittleren Unternehmen (KMU) beitragen kann. Die Forschungs- und Entwicklungsarbeiten hatten zum Ziel, eine Empfehlung zu geben, welche Parametereinstellungen an den analysierten Maschinen vorzunehmen sind und welche potentielle Energieeinsparung diese Maßnahmen mit sich bringen können. Dafür wird die optimale Machine Learning-Methode für dieses Problem ermittelt. Durch die Verwendung des Random Forest-Algorithmus und darauf aufbauend des künstlichen neuronalen Netzes werden schließlich Zusammenhänge zwischen Parametereinstellungen an den Maschinen und Einsparung von Energie sowie Verbesserung der Qualität der produzierten Teile aufgezeigt und in Modellen dargestellt. Das durchgeführte Projekt konnte aufzeigen, dass KMU mit dem Potential von Verfahren der künstlichen Intelligenz in Verbindung mit Digitalisierungsmaßnahmen in der Produktion und der Datenerfassung sowohl erhebliche ökologische als auch wirtschaftliche Vorteile hervorbringen können.

1 Problemstellungen und Hauptzielstellung

In den letzten Jahren hat die Bedeutung des betrieblichen Umweltschutzes in Unternehmen immer mehr zugenommen. Es gelten vielfältige gesetzliche Bestimmungen, die es zwingend erforderlich machen, den Umweltschutz nicht nebenbei zu betreiben, sondern ihn in Kerngeschäftsprozesse zu integrieren, um einen wirtschaftlichen Erfolg zu sichern¹. Der betriebliche Umweltschutz hat einen sehr hohen Datenbedarf, dessen Deckung vor allem für kleine und mittlere Unternehmen (KMU) sehr aufwändig und dadurch teuer ist. Um der Berichtspflicht gegenüber dem Gesetzgeber nachzukommen oder die internen Prozesse ökologisch bzw. ökonomisch zu optimieren, müssen verschiedenste Daten erfasst und aufbereitet werden.

Bedingt durch die stark dynamischen Entwicklungsprozesse (ökonomische, demographische, soziologische, ökologisch-biologische Prozesse) ergeben sich hohe Anforderungen an das Auffinden von wissenschaftlich fundierten Antworten auf umweltspezifische Fragestellungen und zur Formulierung von Modellen hinsichtlich der Verbesserung der Effizienz des Materialeinsatzes, der Verringerung von Energie und Abfall sowie der Minderung des CO₂-Ausstosses. Entscheidungsfindungen werden damit einerseits durch die Kürze der zur Herbeiführung von Lösungen verfügbaren Zeit und andererseits durch die Komplexität der Systeme und Prozesse des Umweltbereiches erschwert. Bei Untersuchungen von Umweltsystemen bildet die große Menge von Daten eine weitere Schwierigkeit. Sie sind in der Regel raum- und zeitabhängig und werden in Geoinformationssystemen, Umweltinformationssystemen oder Umweltdatenbanken gesammelt und gespeichert. Die sinnvolle Kopplung von Modellen oder auch Simulationstechniken für den betrieblichen Umweltschutz und Umweltbezug ist heute noch Gegenstand intensiver Forschungsarbeiten².

Ziel des Projektes Nova bestand in der Optimierung der Anlagen hinsichtlich des Ressourcenverbrauchs und des Abfallaufkommens. Dafür wurden die Maschinendaten erfasst sowie analysiert, um Optimierungspotenziale der Stoff- und Energieströme aufzudecken und Verbesserungen bezüglich Materialien, Abfall und Energie umzusetzen. Besonderer Fokus dabei lag in der Minderung des CO₂-Ausstoßes. Zur Erfassung der notwendigen Maschinendaten entwickelte die HTW geeignete Schnittstellen für den Datenimport.

Eine weitere Problemstellung, welche im Projekt bearbeitet wurde, ist die Analyse der Fertigungsprozesse, welche sich durch starke Nichtlinearitäten und zeitinvariantes Verhalten auszeichnen mit Big Data-Verfahren und selbstlernenden Steuerungen. Dafür ist maschinelles Lernen dann geeignet, wenn ausreichend Trainingsdaten zur Verfügung stehen. Aufgrund des hohen Materialdurchsatzes bei der Fertigung der Kunststoffteile bei NOVAPAX und anderen KMU im Kunststoffsektor wurden die Voraussetzungen für die Entwicklung geeigneter Lernverfahren erfüllt. Dabei wurden die Daten nicht nur offline, sondern bereits im laufenden Betrieb trainiert. Daneben sollten auch indirekte Faktoren (z.B. Wartungsintervalle, mechanische Störungen) berücksichtigt werden. Als Antwort auf die zunehmende Bedeutung der heutigen Informationstechnologien in industriellen Produktionsprozessen strebte das Projekt an, diese Technologien zur nachhaltigen Digitalisierung einzusetzen, um die Umweltauswirkungen der Industrie zu verringern und die Effizienz zu erhöhen.

¹ Vgl. Baumast A., Pape, J. (2009): Betriebliches Umweltmanagement - Nachhaltiges Wirtschaften in Unternehmen, Stuttgart.

² Vgl. Arndt, H.-K.; Gómez, J. M.; Wohlgemuth, V.; Lehmann, S.; Pleshkanovska, R. (Hrsg.): „Nachhaltige Betriebliche Umweltinformationssysteme: Konferenzband zu den 9. BUIS-Tagen“, Springer Gabler, Januar 2018, ISBN-13: 978-3658203795, S. 97 ff.

Ziel dieses Ansatzes war es, moderne Software und Informationstechnologie zu nutzen, um die Nachhaltigkeit in verschiedenen interdisziplinären Anwendungsbereichen zu verbessern. Gleichzeitig konzentrierte man sich auf die Senkung des Energieverbrauchs und die Steigerung der Energieeffizienz durch Einbeziehung von Umweltaspekten in die Herstellung und den Einsatz von Informationstechnologie. Die Simulation von Stoffströmen spielt bei beiden Ansätzen eine wichtige Rolle. Basierend auf den Kennzahlen der Umweltbilanz 2016 der Fa. NOVAPAX von Wasser, Energie und Abfall konnte eine jährliche Gesamtsumme von ca. 2.500 t CO₂-Emissionen in die Atmosphäre ermittelt werden. Bei einer Ausschussquote von 10% bei einem Materialeinsatz von 1.400 t/Jahr produziert NOVAPAX 140 t Abfall/Jahr.

Mit der im Projekt eingesetzter Umweltmanagementsoftware und modernen Analyse- und Auswertungsmethoden wurde eine Prozessoptimierung bezüglich der Vermeidung von Losgrößen mit kleinen Stückzahlen, Verringerung von Qualitätsfehlern und Steigerung des gewinnbringenden Recyclings von Wertstoffen (z.B. Polyoxymethylen) durch intelligente Angebotsbetrachtung erreicht, welche die geplanten Umweltziele realisierbar macht und die praktische Umsetzung langfristig und nachhaltig ermöglicht.

2 Wissenschaftlich-technische Ergebnisse

In diesem Abschnitt werden die erreichten wissenschaftlich-technischen Ergebnisse präsentiert.

2.1 Datenvorverarbeitung und -analyse

Im ersten Schritt wurden die erforderlichen Daten innerhalb des Unternehmens aus verschiedenen Softwarelösungen in einer für das Projekt entwickelten SQL-Datenbank zusammengefasst, validiert und verifiziert. Die Datenaufbereitung wurde mit Hilfe der Frameworks Pandas und Numpy durchgeführt. Hierzu zählen Datenbereinigung, Normierung bzw. Standardisierung und Dimensionsreduktion. Mögliche Ansätze für das maschinelle Lernen sind überwachtes Lernen und unüberwachtes Lernen. Für das überwachte Lernen ist eine Klassifikation sinnvoll, die mittels dem Framework Scikitlearn umsetzbar ist. Für das unüberwachte Lernen bietet Clustering hohe Erfolgsaussichten, da durch die Diversität der Daten eine Bearbeitung losgelöst von Gruppenzugehörigkeiten sinnvoll erscheint. Für die Modellauswahl bietet das Framework TensorFlow viele Algorithmen an, alle Frameworks sind kompatibel mit Python.

Die Maschinen bei Novapax protokollieren Daten über jedes produzierte Teil auf einer internen Festplatte. Diese sind über das sog. OPC-Protokoll abrufbar. Ein Computer im firmeninternen Netzwerk kopiert diese periodisch in eine MS-SQL Datenbank. Mittels einer VPN-Verbindung ist die Datenbank auch extern erreichbar. Um sowohl die Geschwindigkeit der Algorithmen zu steigern als auch die Netzwerkinfrastruktur des KMUs zu schonen, wurde die Datenbank auf den lokalen Computer kopiert, auf dem die Berechnungen durchgeführt werden.

Die Daten sind in drei Tabellen unterteilt:

- Produktionsprotokoll: Produktionsparameter
- Alarmprotokoll: Störungen bei der Produktion
- Einrichtprotokoll: Werkzeugwechsel

Merkmal	Ein-	Beschreibung
Maschinen_Zeit-	date	Datum und Zeit des Schusses
Schuss_ID	int	ID des Schusses (DB Zähler)
Qualität	int	Maschineninterner Qualitätsindikator (1 Gutteil, 3 Schlechtteil)
Schuss_nach_Neustar	int	Maschineninterner Zähler
Zykluszeit	sec	Dauer eines Zykluses
Massepolster	cm ³	Ausgleichsmasse für Abkühlung (Teil zieht sich zusammen)
Umschaltvolumen	cm ³	Ab diesem Volumenwert wird der Prozess umgeschaltet (Nachdrücken)
Umschaltspritzdruck	bar	Druck für Nachdrücken
Maximaler_Spritz-	bar	Maximaler Druck des Prozesses
Einspritzzeit	sec	Dauer des ersten Prozessschrittes (Ausfüllen der Form)
Dosierzeit	sec	Dauer des Schmelzvorganges für den nächsten Prozess
Zylinderheizzone	°C	Heizzone für Schmelzung des Granulats
Werkzeugkreis	°C	Steuert die Temperatur des Werkzeuges
Temperi-	°C	Temperieren (Wasser) des Werkzeugkreises
Integral	<i>p/s</i>	Von Druck über Zeit
Energiezähler	<i>W/h</i>	Zählerstand

Tabelle 1 Produktionsprotokoll

Merkmal	Einheit	Beschreibung
ID	int	Datenbank-ID
Maschinen_Zeit-	date,	Datum und Zeit des Schusses
Alarm_ID	string	Fehlercode des Alarms
Alarm	string	Beschreibung des Alarms
Node_betroffener_ Parameter	int	Auflistung betroffener Nodes
auslösendes_Sys-	int	ID des auslösenden Systems
Nachrichten_Kate-	int	Kategorie der Nachricht
Nachrichten_Typ	int	Typ der Nachricht

Tabelle 2 Alarmprotokoll

Merkmal	Einheit	Beschreibung
ID	int	Datenbank-ID
Maschinen_Zeit-	date,	Datum und Zeit des Schusses
Start_Parameter_ betroffen	bool- ean	ID des Schusses (DB Zähler)
Änderung	string	Beschreibung der vorgenommenen Änder-
Benutzer	string	Durchführender Benutzer
betroffene_Nodes	string	Auflistung der betroffenen Nodes
Nachrichten_Typ	int	Typ der Nachricht

Tabelle 3 Werkzeugwechsel

Die Zykluszeit beträgt ca. 25 Sekunden, was bei einem optimalen Produktionstag theoretisch 3.456 Schüsse (produzierte Teile) ergibt. Jeder Schuss ist in der Produktionsdatenbank gespeichert und hat eine Größe von 164 Byte, somit ergeben sich bei optimaler Auslastung ca. 0,56 Megabyte Daten am Tag. Der Fokus hierbei lag auf dem Feld *Qualität* aus der Produktionsdatenbank, da dieses Feld den Ausschuss beschreibt. Mögliche Korrelationen bei "Qualität 3" (Schlechtteil) mit einem Alarm aus dem Alarmprotokoll oder einer Änderung der Maschinenparameter aus dem Einrichtprotokoll sowie Abweichungen bei den Produktionsparametern sind denkbar.

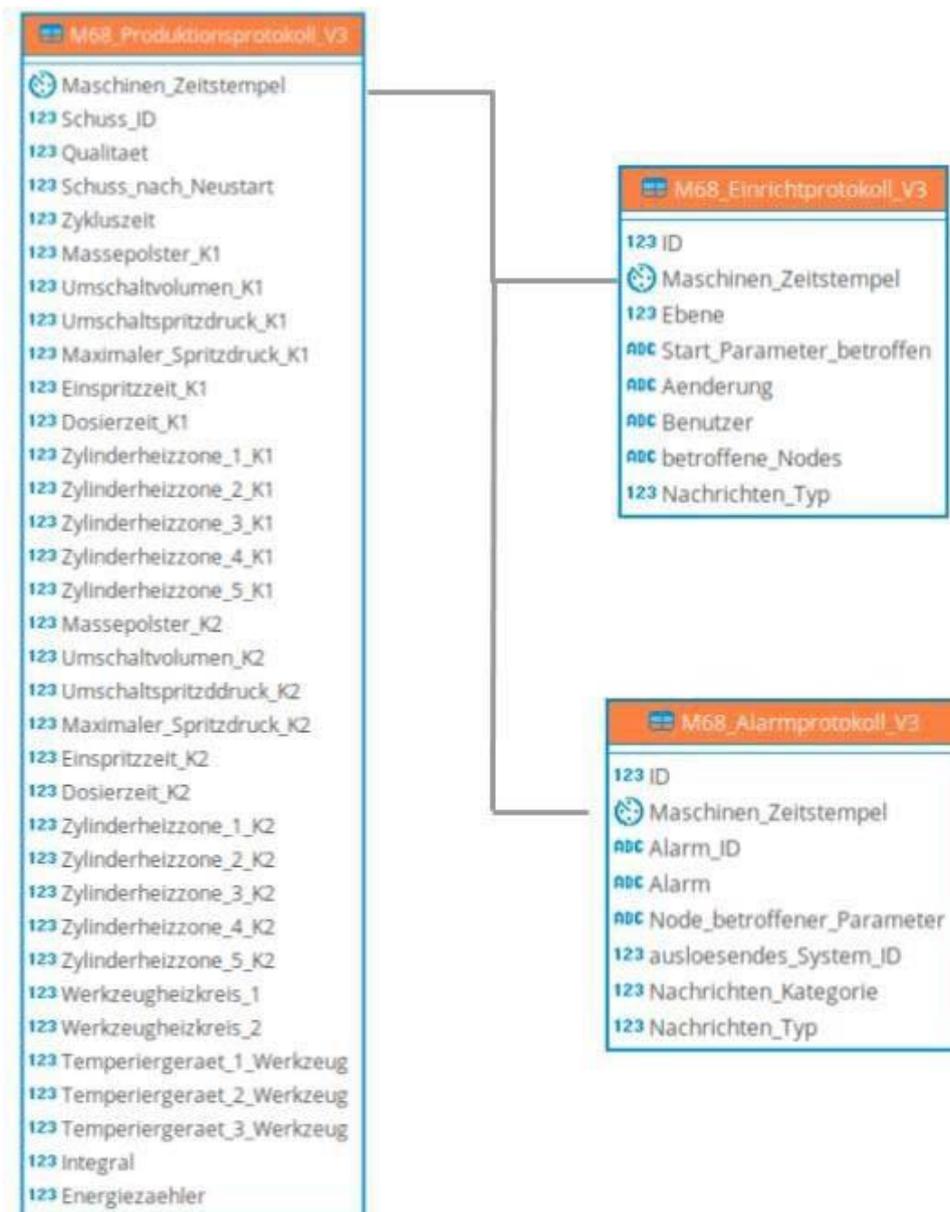


Abbildung 1 Beziehungsschemata der Datenbank

Die Beziehung zwischen den Datenbanken ist durch das Feld `Maschinen_Zeitstempel` gewährleistet. Es stellt einen gemeinsamen Zeitmaßstab her, wodurch sich Ereignisse und Zyklen vergleichen lassen. Weitere interessante Werte sind die Felder `“Zykluszeit”` - loggt die Dauer eines Zyklus - `“Alarm”` - speichert die Alarmpfeilungen - `“Integral”` - Druck über Zeit des Prozesses – und `“Energiezaehler”` - zählt die verbrauchten Kilowattstunden. Daten zu der nachgelagerten optischen Qualitätsprüfung fehlen in der hier aktuellen Version der Datenbank. Dies schränkte die Aussage-

kraft eines auf dieser Datenbasis erstellten Modells erheblich ein, da die von der Maschinenüberwachung nicht erkannten Fehlteile nicht untersucht werden konnten.

Die Datensätze wurden auf Duplikate und fehlende Werte untersucht. Je nach Eigenschaften der Merkmale mussten dann die Attribute standardisiert, normiert oder skaliert werden. Falls die Merkmale Gleitkommazahlen beinhalten, wurden diese auf Ganzzahlen (Integer) skaliert. Eine erste Analyse der Daten zeigte, dass die *Einspritzzeit_K2*, *Umschaltvolumen_K2*, *Massepolster_K2*, *Maximaler_Spritzdruck_K1*, *Umschaltspritzdruck_K1* und *Massepolster_K1* die größten Auswirkungen auf die Varianz der Qualität hatten. Attribute, welche einen hohen Einfluss auf die Varianz der Qualität haben, sollten ausgewählt werden, um den Algorithmus zu trainieren. Um Über- bzw. Unteranpassung zu vermeiden, wurde eine verkleinerte Auswahl der Trainingsmerkmale benutzt.

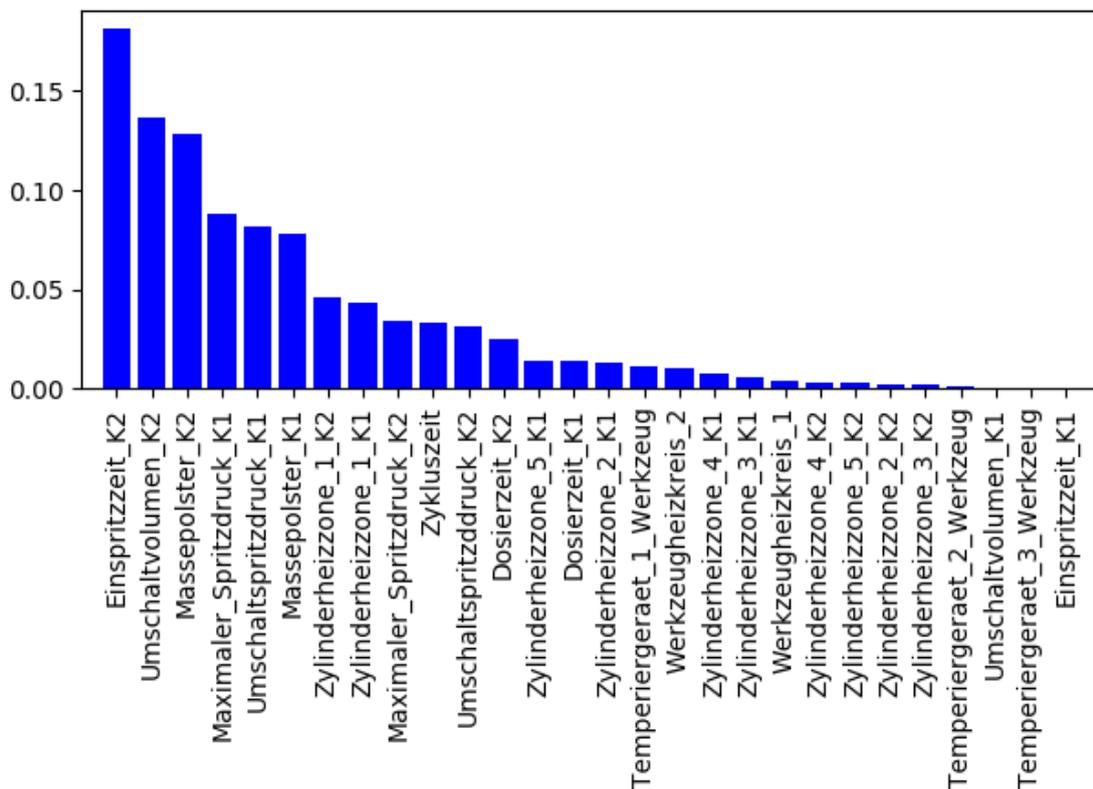


Abbildung 2 Bedeutung der Merkmale in Bezug auf Qualität

Für das geplante Training mit unüberwachtem Lernen oder Clustering muss keine spezielle Auswahl getroffen werden, da der Lernalgorithmus keine vordefinierten Zielvariablen benötigt.

Die Aufteilung der Datensammlung wurde mit einem Befehl aus dem Scikitlearn Framework durchgeführt: Die Datensätze x, y wurden zufällig mit dem Verhältnis 70:30 ($test_size=0.3$), in $train$ und $test$ Datensätze aufgeteilt. Der Parameter $random_state$ legt die Aufteilung fest und diente dazu, die Ergebnisse zu reproduzieren.

```
In [4]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
#Aufteilen der DataFrames in Trainings und Testdaten
#test_size: Verhältnis von Trainings- zu Testdaten (hier 70:30)
#random_state: Zufällige Verteilung der Daten, Wert gibt Parameter zu Wiederholbarkeit an
X_train.head()

Out[4]:
```

	Zykluszeit	Massepolster_K1	Umschaltvolumen_K1	Umschaltspritzdruck_K1	Maximaler_Spritzdruck_K1	Einspritzzeit_K1	Dosierzeit_K1	Zylinderheizz
1120	31.008	4.542247	12.99987	903.6863	1926.417	1.63	4.284	
26777	32.584	4.578215	12.99987	923.6367	1967.962	1.63	4.268	
28579	32.544	4.480587	12.99987	918.2654	1964.564	1.63	4.272	
5184	30.960	4.629598	12.99987	911.1403	1925.431	1.63	4.292	
10110	30.982	4.578215	12.99473	903.4670	1921.265	1.63	4.292	

5 rows x 28 columns

Abbildung 3 Aufteilung der Dataframes

In einem weiteren Schritt wurde die Anzahl der Datenbestände für die jeweilige Maschine per SQL-Abfrage betrachtet. Zusätzlich wurde abgefragt, wie viele Werte dabei von fehlerfreien (Qualitaet = 1) und wie viele von fehlproduzierten Teilen (Qualitaet = 3) stammen.

Maschine	Anzahl Datensätze	Gutteile		Schlechtteile	
		Absolut	Prozent	Absolut	Prozent
M64	168.917	166.232	98,4	2.685	1,6
M67	247.052	245.982	99,6	1.070	0,4
M68	86.203	85.325	99,0	878	1,0
M69	275.286	270.780	98,4	4.506	1,6
Gesamt	777.458	768.319	98,8	9.139	1,2

Abbildung 4 Datensätze der betrachteten Maschinen und deren Anteile von Gut- und Schlechteilen

Der Erfassungszeitraum der Daten lag zwischen dem 18.03.2019, 12:56:28 und 08.07.2019, 16:47:20. Für die spätere Energieberechnung (siehe Abschnitt 4.5) ist eine kontinuierliche Datenerfassung optimal. Zur Analyse zeigt Abbildung 4.1, wann das Skript Daten erfasst hatte. Es existieren einige Lücken im Erfassungszeitraum. Hier musste eine differenzierte Betrachtungsweise an den Tag gelegt werden, denn einige Maschinen produzieren nicht 24 Stunden am Tag und daher entstanden natürlicherweise scheinbare Unterbrechungen in den Daten. Das erschwerte die Vorverarbeitung zur Berechnung der Energie.

Bei der Betrachtung der Datenvollständigkeit war auffällig, dass bei allen Maschinen komplett weiße Einträge vorlagen, welche das Fehlen von Daten bedeuten. Dabei handelte es sich nicht um fehlende Daten, sondern den Umstand, dass durch die unterschiedlichen Bauweisen diese Messwerte nicht erhoben wurden, da beispielsweise die Zylinderheizzone nicht existiert. Im Allgemeinen lag ein vollständiger Datensatz vor. Auffällig waren die Lücken bei Zylinderheizzone_K2_[3,4,5] der Maschine M69. Durch die zusammenhängenden, fehlenden Daten war davon auszugehen, dass in diesem Fall die Produktion der Maschine umgestellt wurde, bei der diese Bereiche nicht genutzt wurden.

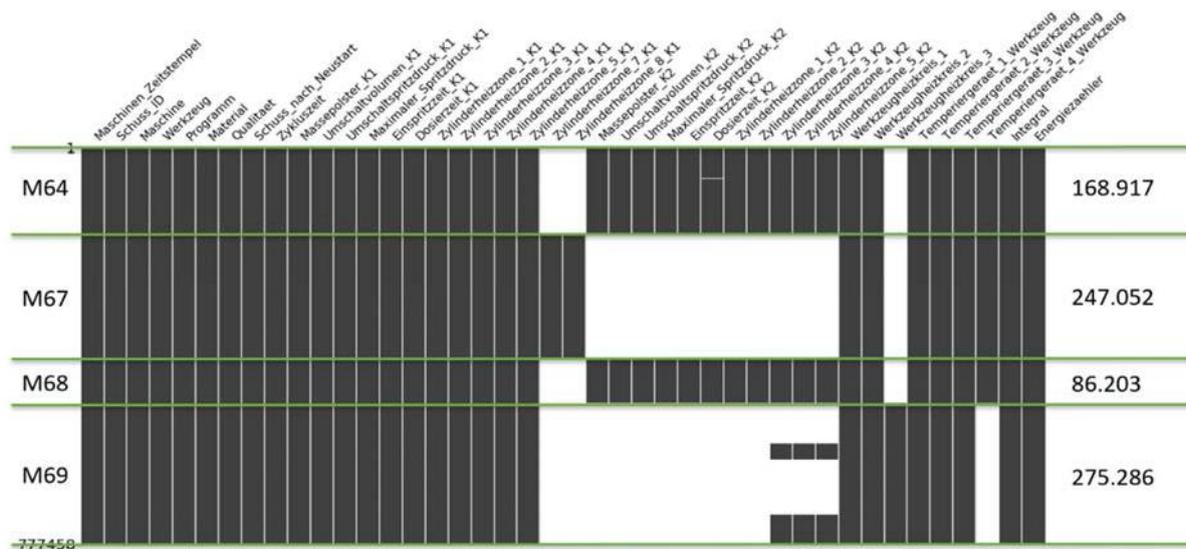


Abbildung 5 Datenvollständigkeit: Schwarze Markierungen stehen für vorhandene, weiße für nichteingetragene Daten

In der Datenbank befinden sich Spalten, wie das Werkzeug, das verwendete Material und das eingestellte Programm. Für eine Untersuchung konnten diese Angaben zwar relevant sein, sie enthalten jedoch durchgängig den gleichen Wert. Daher liefern sie keinen Informationsgewinn. Bei der Modellerstellung wurden sie aus den Daten entfernt, um die Berechnungsgeschwindigkeit des Modells zu erhöhen.

2.2 Modellgenerierung und -analyse

Im nächsten Schritt wurde bestimmt, welcher Algorithmus am besten für die vorliegende Problemstellung geeignet ist. Zunächst wurde eine Einordnung in die drei Arten des Machine Learnings vorgenommen. Das unüberwachte Lernen verarbeitet Daten ohne Labels. In diesem Fall sind jedoch die Einträge bereits kategorisiert. Damit wird eine bereits bekannte Information ignoriert, wodurch ein ungenaueres Ergebnis zu erwarten ist. Zudem ist das Ziel bereits eindeutig definiert: Die Reduzierung der Fehlproduktionen und Senkung des Energieverbrauchs. Eine Clustering-Analyse könnte unter Umständen Gruppen bilden, die zwar ähnliche Eigenschaften besitzen, aber nicht zur Beantwortung der Forschungsfrage beitragen. Mit einer geeigneten Interpretation ließen sich unter Umständen auch Ergebnisse erzielen, durch Einbeziehung aller vorhandenen Informationen, insbesondere der Labels, versprechen die anderen Arten des Machine Learnings jedoch mehr Erfolg. Die Anwendung von verstärkendem Lernen ist eine mögliche Art Ergebnisse zu liefern. Dafür war allerdings eine Grundvoraussetzung nicht gegeben: Die Interaktion mit der Umgebung. Im Projekt lagen bereits erfasste Daten vor. Um diese Methode zu verwenden, müsste der Algorithmus direkte Veränderungen an der Maschine vornehmen und mit den Ergebnissen die Parameter anpassen. Auch wenn so in der Theorie vielversprechende Ergebnisse zu Tage geliefert werden könnten, war der

finanzielle Aufwand im Vergleich zu den möglichen Einsparungen durch die Forschungserkenntnisse zu groß.

Die Voraussetzungen für überwachtes Lernen waren in diesem Fall vollständig erfüllt: Es existierten vorgelabelte Daten und es war definiert, in welcher Weise diese Informationen analysiert und optimiert werden sollen. Basierend darauf konnten Modelle erstellt und auf Genauigkeit überprüft werden. Konsequenterweise wurde daher die Methode des überwachten Lernens eingesetzt. Bevor einzelne Algorithmen bewertet wurden, galt es zunächst, die Anforderungen zu beschreiben. Mit etwa 600.000 Datensätzen lag eine große Menge an Informationen vor, die verarbeitet werden mussten. Der Algorithmus musste in der Lage sein, mit dieser Größe umzugehen. Über das Modell konnten keine Annahmen gemacht werden, daher sollte bei dessen Erstellung auch nicht-lineare Zusammenhänge erkannt werden. Die Problemstellung erforderte, dass eine Methode nicht nur korrekte Vorhersagen macht, sondern auch interpretierbar ist und die Gründe für eine Entscheidung offenlegt. Im besten Fall sollte die Erstellung des Modells zeitlich möglichst kurz sein, da der Zeitraum der Arbeit begrenzt ist und mehrfache Kalkulationen über mehrere Tage schnell die Erfüllung der Aufgaben behindern können. Die Zielsetzung beinhaltete daher die Erklärung der Qualität in Form einer diskreten Klassifizierung sowie eine kontinuierliche Abschätzung in Form einer Regression für den Energieverbrauch. Optimalerweise sollte der Algorithmus beide Arten beherrschen, um den Entwicklungsaufwand zu reduzieren. Außerdem wäre ein geringes Risiko einer Überanpassung von Vorteil, um ein allgemeines Modell zu erhalten. Sowohl der Random Forest-Algorithmus als auch die künstlichen neuronalen Netze sind für das vorliegende Szenario anwendbar. In Anbetracht der Geschwindigkeit der Modellerstellung und einem einfacheren Einblick in die Entscheidungskriterien für Klassifikationen, bietet der Random Forest mehr Vorteile und wurde daher im ersten Schritt der Umsetzung verwendet.

Zuerst wurde eine Überlegung angestellt, welche Parameter von der Analyse ausgenommen werden sollen. Die Schuss_ID ist eine Referenz aus der Datenbank und enthält keine Information der Maschine. Der Ausschluss dieser Spalte verhinderte, dass der Algorithmus sich einzelne Datensätze anhand der ID merken kann. Ähnlich verhalten sich die geloggte Uhrzeit und das Datum in der Spalte Maschinen_Zeitstempel. Auch die von der Schuss_ID abgeleitete Größe Schuss_nach_Neustart wurde vernachlässigt. Für spätere Aussagen zu Änderungen der Maschineneinstellungen sind diese Parameter nicht zielführend. Die Spalten Programm, Werkzeug und Material enthalte für jede Maschine nur einen Wert, was diese Spalten überflüssig macht. Wie bereits beschrieben, besitzt nicht jede Maschine für alle Spalten valide Werte. Datenfelder, die nur Nullen enthalten, wurden ebenfalls gefiltert.

Anschließend wurde der Datensatz in Trainings- und Testdaten unterteilt. Als Testdatengröße wird 25% gewählt. Der Algorithmus unterteilte die Daten nach dem Zufallsprinzip.

Wichtig ist das Definieren des Parameters random_state. Zufallszahlen werden meist durch einen deterministischen Algorithmus generiert. Diese nennt man Pseudozufallszahlen, da sie zwar statistisch, durch die Vorhersagbarkeit jedoch nicht tatsächlich zufällig sind. Durch die Kenntnis einer

generierten Zahl in der Folge lassen sich alle anderen Werte berechnen³. Jeder Generator benötigt einen Startwert, den sog. Seed, der in diesem Fall mit dem `random_state` gesetzt wurde. Die Vorhersagbarkeit ist in diesem Umfeld vorteilhaft. Denn so konnte garantiert werden, dass bei jedem Start des Programms dieselbe Unterteilung vorgenommen wurde. Somit sind spätere Hyperparameteranpassungen vergleichbar. Andernfalls könnten Änderungen im Ergebnis auch auf den Umstand zurückführbar sein, dass andere Unterteilungen stattgefunden hatten. Für den ersten Testdurchlauf wurden Standardwerte verwendet, um zunächst die Funktion des Skriptes zu verifizieren. Im Anschluss wurde das Modell mit den Trainingsdaten erstellt. Danach sagte der Algorithmus die Labels der Testdaten voraus. Die Funktion `accuracy_score` verglich die Vorhersage mit den richtigen Ergebnissen und gab die prozentuale Erkennungsrate zurück.

Im nächsten Schritt musste das Modell für eine optimale Erkennung angepasst werden. Dazu galt es, zunächst herauszufinden, welches eine geeignete Evaluationsmethode für ein gutes Modell ist. Das zuvor berechnete Ergebnis betrug 99,2%. Dies steht für den Anteil an korrekt klassifizierten Teilen. Auf den ersten Blick scheint diese Zahl ein gutes Ergebnis zu sein. Der Wert ist aber abhängig davon, wie die Verteilung der Eingabedaten ist. In diesem Fall lag der Anteil an Gutteilen bei 98,4%. Ein Algorithmus, der immer ein Gutteil vorhersagt, hätte demnach eine Klassifizierungsgenauigkeit von 98,4%. Mittels der Konfusionsmatrix wurde beschrieben, wie viele Gutteile und wie viele Schlechteile richtig erkannt wurden.

	Gutteil	Schlechtteil
Klassifizierung Gutteil	41.510	188
Klassifizierung Schlechtteil	9	471

Tabelle 4 Klassifizierung der produzierten Teile in Abhängigkeit des Labels

Mit diesem detaillierten Einblick in das Ergebnis des Modells ließen sich bessere Aussagen über dessen Qualität machen. Während die Gutteile fast zu 100% richtig erkannt wurden, liegen die Schlechteile mit etwa 71% Erkennungsrate in noch keinem akzeptablen Bereich. Es wurde ein Ziel von 90% angestrebt. In dieser Spanne ist anzunehmen, dass keine Überanpassung stattgefunden und das Modell eine Verallgemeinerung der Datensätze vollzogen hatte. Die Anzahl der Datensätze von etwa 42.000 waren die gewählten 25% Testanteil von der Gesamtheit der Messdaten. In weiteren Versuchen wurden testweise einige Hyperparameter angepasst. Erhöht man die Anzahl der Bäume auf 150, sank die Erkennungsrate sogar auf 72 % bei Schlechteilen, während die Gutteile weiterhin mit nahezu 100 % korrekt klassifiziert wurden. Ein Anheben der Baumtiefe von 6 auf 10 mit den sonst gleichen Parametern führte zu einer Klassifizierungsrate von 41.510/41.519 ($\approx 100\%$) für Gutteile und 596/711 (83,4%) für Schlechteile. Damit ist ein Anstieg zu erkennen. Mit der erhöhten Tiefe ist allerdings auch ein indirektes Speichern der Datensätze wahrscheinlicher. Ein Baum

³ Love, R. Linux-Kernel-Handbuch: Leitfaden zu Design und Implementierung von Kernel 2.6. Open source library. Addison-Wesley, 2005. ISBN: 9783827322470

mit so vielen Blättern wie Datensätzen könnte für jede Kombination einen eigenen Pfad kreieren. In einem weiteren Test wurde das Undersampling evaluiert: Bisher lag das Verhältnis zwischen fehlerfreien und fehlerbehafteten Daten bei etwa 99:1. Optimalerweise wäre jedoch eine Zusammensetzung von 50:50. In diesem Versuch wurden daher alle Schlechteile verwendet und die gleiche Anzahl an Gutteilen zufällig aus der Datenbank ausgelesen. Nun wurde erneut ein Random Forest-Modell erzeugt. Bei diesem Modell ist davon auszugehen, dass ein Overfitting stattfand, da die Klassifizierungsrate sehr hoch war.

	Gutteil	Schlechtteil
Klassifizierung Gutteil	656	4
Klassifizierung Schlechtteil	3	655

Tabelle 5 Klassifizierung der produzierten Teile in Abhängigkeit des Labels (Undersampling)

Nachdem nun stichprobenartig einige Hyperparameterkombinationen getestet wurden, wurde eine Methode entwickelt, um die optimalen Hyperparameter strukturiert festzulegen. Da jede Parameterkombination ca. 5 Minuten dauert, wurde eine Stapelverarbeitung implementiert, die verschiedene Anordnungen testet. Wichtig dabei war, dass eine Vergleichbarkeit der Einstellungen vorliegt. Dies gilt insbesondere für die Festlegung der gewählten Datensätze.

Die Wichtigkeit der Merkmale ergab einen ersten Hinweis auf die Entscheidungskriterien des Algorithmus. Aber die alleinige Kenntnis, welches Feature entscheidend ist, genügt noch nicht dem Anspruch der Problemstellung. Es soll eine genaue Angabe des Bereichs erfolgen, in dem die beste Leistung der Maschinen erzielt wird, damit der Ausschuss verringert, weniger Plastik produziert und weniger CO₂ Emissionen ausgestoßen werden.

Eine zielführende Methodik ist die sog. Brute-force-Methode. Dabei werden möglichst viele Kombinationslösungen ausprobiert⁴. Im Zusammenhang mit diesem Projekt bedeutete dies, Datensätze in unterschiedlichen Kombinationen zu generieren und mit dem Modell zu testen. Für jedes Feature wurde aus der Datenbank der Minimal- und Maximalwert bestimmt. Da kontinuierliche Werte vorlagen, konnte nicht die Gesamtheit aller Möglichkeiten generiert werden. Für diese Überlegung wurde zunächst eine konservative Anzahl von 10 Intervallen gewählt. Im Anschluss wurde analysiert, an welchen Stellen die Qualität von „gut“ auf „schlecht“ wechselt. Diese Methodik erfordert jedoch eine sehr lange Rechenzeit. Bei Betrachtung aller Parameter und ihrer Kombination wäre man auf eine Zeitdauer von 31 Mrd. Jahren gekommen. Daher kam die lokale Brute-force-Methode mit LIME zum Einsatz. Attraktiv an der Brute-force-Methode ist ihre Gründlichkeit. Es ist möglich, sie angepasst mithilfe der Methode der local interpretable model-agnostic explanations

⁴ Mukkamala, K. Hands-On Data Structures and Algorithms with JavaScript: Write efficient code that is highly performant, scalable, and easily testable using JavaScript. Packt Publishing, 2018. ISBN: 9781788397544.

(LIME)⁵ einzusetzen. Anstatt zu versuchen, ein globales Modell zu erstellen, kann diese Methode für die Umgebung eines bestimmten Datensatzes Erklärungen liefern. Dazu generiert das Verfahren Stichproben zufällig, aber in gleichmäßigen Abständen im Umfeld, und gewichtet diese je nach Entfernung zum ursprünglichen Punkt. Daraus liefert der Algorithmus ein lineares Erklärungsmodell.

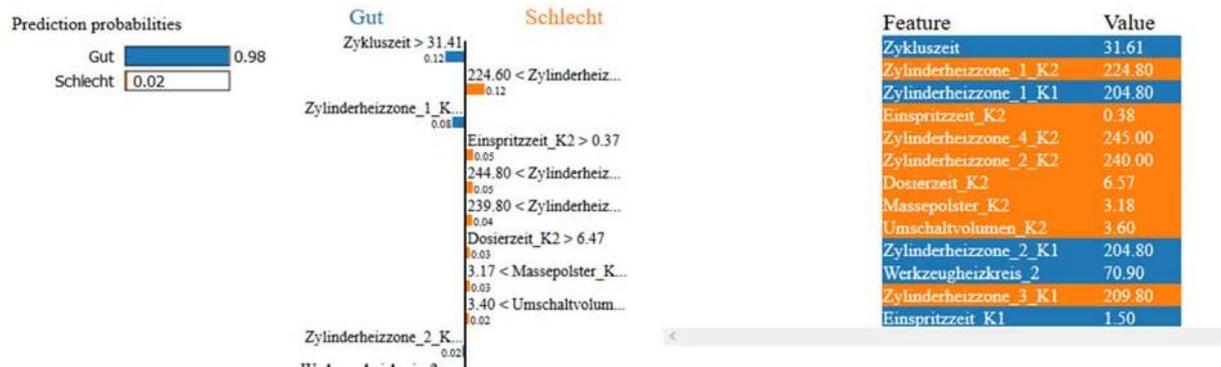


Abbildung 6 Erklärungsmodell der LIME-Methode

Links stehen die Vorhersagen des Random Forest, rechts die Werte, die die einzelnen Features haben, und die Mitte zeigt die Erklärungen von LIME. Die LIME-Methodik wurde im nächsten Schritt auf alle Testdatensätze erweitert. Die Klassifizierungen für jedes Feature eines Datensatzes wurden nun gruppiert und geplottet.

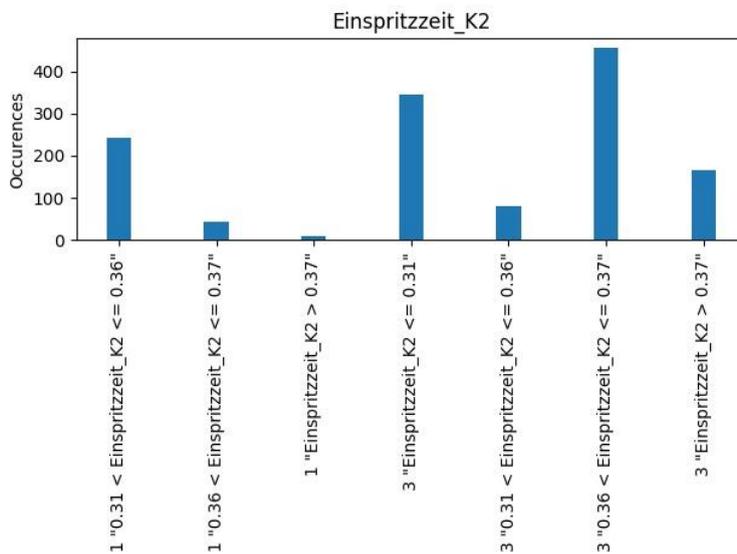


Abbildung 7 Ergebnis von addierten lokalen Approximationen (1: gute Qualität, 3: schlechte Qualität)

⁵ Tulio Ribeiro, M., Singh, S. und Guestrin, C. "Why Should {I} Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd {ACM} {SIGKDD} International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016. 2016, S. 1135-1144.

Mit dieser Grafik ließen sich schon genauere und allgemeinere Aussagen treffen. Auffällig ist zunächst, dass deutlich mehr Schlecht- als Gutteile in der Grafik aufgelistet sind. Der Grund dafür ist die Sortierung nach Wichtigkeit. Betrachtet man die relativen Größen der Grafik innerhalb einer Klassifizierung, dann stellt man fest, dass die gute Qualität zwischen 0.31 und 0.36 ein deutliches Maximum zeigt. Umgekehrt ist ein deutlicher Abfall an dieser Stelle bei schlechter Qualität zu sehen, wohingegen außerhalb dieses Bereichs eine hohe Rate vorliegt.

Somit lässt sich erkennen, dass bestimmte Bereiche für ein Feature relevant sind. Mit der LIME-Methode sind allerdings keine detaillierten quantitativen Aussagen möglich. Die aufaddierten Punkte wurden alle gleich gewichtet, unabhängig davon, ob an dieser Stelle eine Wichtigkeit von 10% oder 0.1% einhergeht. Auch lassen sich die Features untereinander schlecht einschätzen: An welchen Merkmalen sollten vorrangig Änderungen vorgenommen werden? Für die Einspritzzeit ist die Genauigkeit bezüglich des Bereichs bereits auf einem hohen Level, für andere Werte ist die Spanne jedoch sehr groß und zu ungenau, um qualitativ gute Aussagen machen zu können. Zusammenfassend war diese Methode ein guter Anfang und zeigte erste Einblicke in die Zusammenhänge zwischen den Maschinendaten und der Qualität. Um detailliertere Ergebnisse zu liefern, sind bessere Eingabedaten nötig. Im ersten Ansatz wurde aus den LIME-Erklärungen lediglich der Klassifizierungsbereich entnommen, nicht aber die genaue Angabe, wie groß der Einfluss ist, den dieser auf das Gesamtergebnis hat. Da für jedes Teil nicht nur der Bereich, sondern auch der genaue Wert bekannt ist, gehen mit dieser Art der Verwendung der LIME-Methode Informationen verloren. Für einen zweiten Ansatz wurde daher erneut eine Erklärung mit der LIME-Bibliothek vorgenommen. Zusätzlich zu den Ergebnissen wurde die ID des jeweiligen Datensatzes hinzugefügt. So ließ sich im Nachhinein jede Größe der einzelnen Merkmale nachvollziehen.

Für den Plot wurde nun der Featurewert jedes Datensatzes mit der Wichtigkeit in dieser Spanne gezählt. Es kam vor, dass für den gleichen Wert mehrere Einträge vorhanden waren. Diese wurden im Plot gemittelt angezeigt.

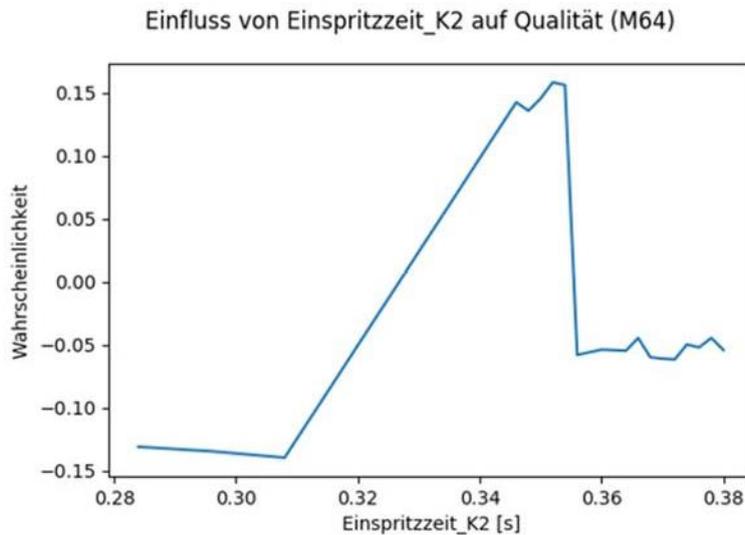


Abbildung 8 Ergebnis von addierten lokalen Approximationen; y-Achse: Wahrscheinlichkeit, mit der der korrespondierende Wert auf der x-Achse zu einer Gutteil-Klassifizierung beiträgt. Negative Werte symbolisieren den Anteil an einer Einstufung als fehlproduziertes Teil.

Das Liniendiagramm ist bei einem geeigneten Datensatz eine gute Darstellungsmethode. Dieser Umstand ändert sich jedoch, wenn Messfehler im Datensatz sind, bei denen vereinzelte Werte fernab vom Durchschnitt liegen. Durch diese wird das Diagramm gestaucht und die eigentlichen relevanten Daten sind nicht ablesbar. Durch das Verschieben der Grenzen des Plots konnte dieser Umstand verbessert werden. Eine weitere Schwierigkeit kann auftreten, wenn Punkte mit stark unterschiedlichen Werten nahe beieinander liegen und somit die Daten nicht mehr interpretierbar sind. Daher wurde ein weiterer Filter eingebaut, der nur Punkte mit entsprechender Anzahl an eingeflossenen Datensätzen zulässt, damit die unwichtigeren Informationen ausgeblendet werden.

Eine allumfassende Aussage für jedes Diagramm zu machen, ist jedoch nicht möglich. Bei einigen Diagrammen existieren mehrere voneinander getrennte Abschnitte mit jeweils vielen Datenpunkten. Ein Filtern von Werten, die zu weit vom Durchschnitt abweichen, kam daher nicht in Frage. Außerdem gibt es Diagramme, die eine hohe Informationskonzentration pro Datenpunkt besitzen, und andere, bei denen selten mehr als 2 Werte in einen Punkt eingehen. Zusammenfassend ist also ein manuelles Bearbeiten für jedes Diagramm nötig. Da eine große Anzahl an Abbildungen zu verarbeiten war, wurde ein Tool zur Verarbeitung programmiert. Dazu wurde zunächst das Programm modifiziert, welches die Grafiken aus den LIME-Erklärungen erstellt. Zusätzlich zu den Abbildungen wurden die berechneten Daten in eine Textdatei geschrieben. In dieser sind der x- und y-Wert sowie die Anzahl an Vorkommnissen des Punktes in den Daten enthalten. Das entwickelte Tool durchsucht die Ergebnisordner nach vorhandenen Textdateien für die jeweiligen Features der Maschine und listet diese auf. Beim Klick auf den Show-Button wird die Auswahl geplottet. Mit den Schiebern im unteren Bereich lässt sich die gewünschte Anpassung zu oberer und unterer Grenze sowie die Anzahl an minimalen Datensätzen pro Punkt einstellen. Durch einen Klick auf das Speichersymbol wird der Plot in eine Datei geschrieben. Mit den so generierten Diagrammen konnte eine qualifizierte

Aussage gemacht werden, an welchen Stellen die jeweiligen Parameter einen Einfluss auf die Qualität haben.

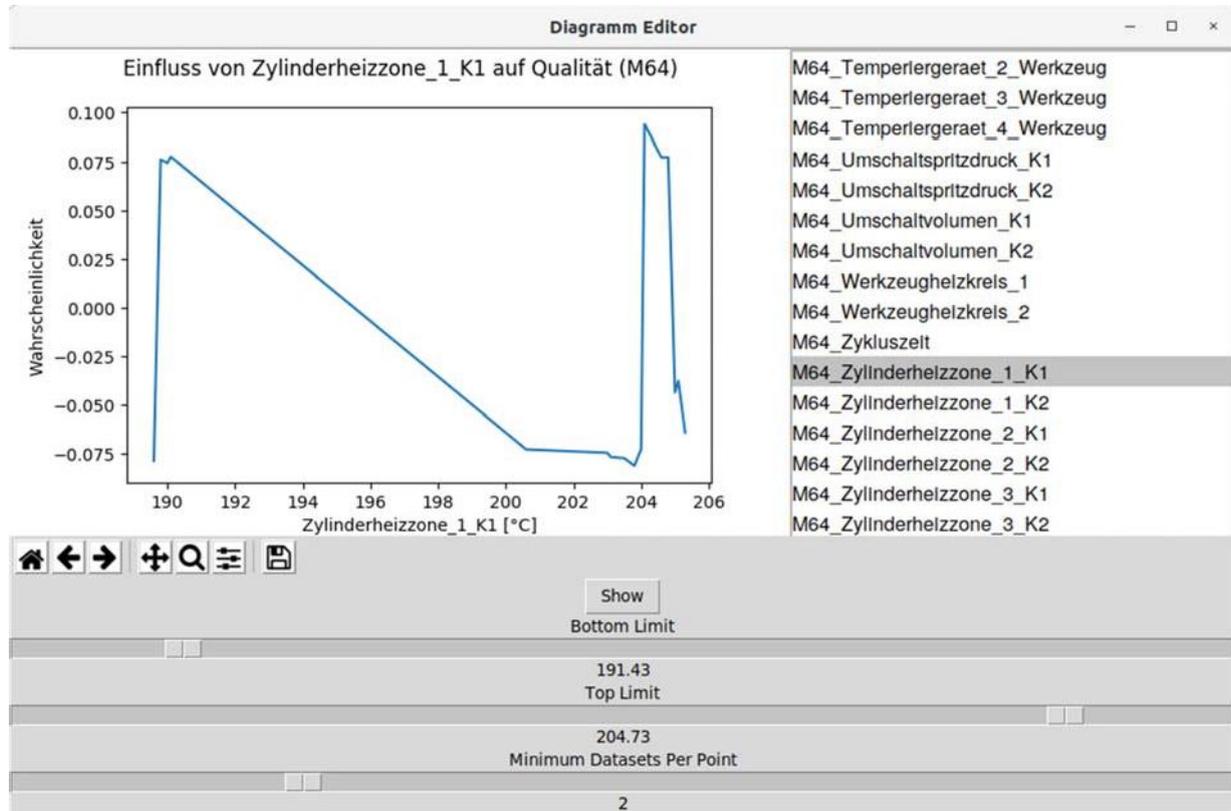


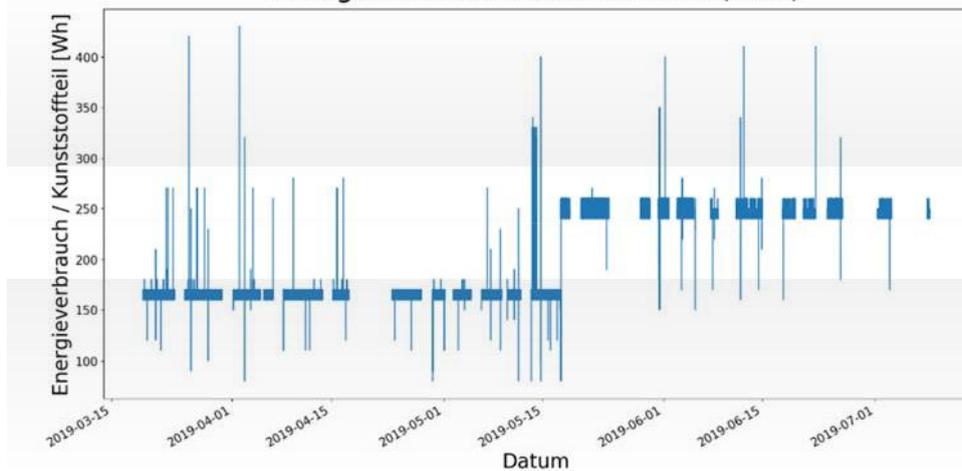
Abbildung 9 Editor zur verbesserten Darstellung der Diagramme

Im nächsten Schritt wurde die beschriebene Vorgehensweise zur Qualitätsanalyse auf die Energieanalyse übertragen. Die zu analysierende Größe war der Energieverbrauch pro produziertem Kunststoffteil. Dazu wurde die Differenz von zwei zeitlich aufeinander folgenden Datensätzen gebildet. Für eine erste Plausibilitätsüberprüfung wurden die so erhaltenen Energieverbräuche pro Zeit geplottet. Daraus war ersichtlich, dass eine weitere Datenfilterung stattfinden muss, bei genauer Betrachtung fiel auf, dass die Spitzen immer nach Zeiträumen auftreten, in denen keine Daten vorhanden sind. Daraus ließ sich schließen, dass zu diesen Zeitpunkten die lückenlose Erfassung aller Teile unterbrochen war. Die Folge in der Berechnung war, dass für das erste Teil nach längerer Pause die Energie berechnet wurde, die für alle Teile im Unterbrechungszeitraum aufgewendet wurde. Zwei mögliche Ursachen für die Lücken in den Datensätzen sind Verbindungsprobleme in der Infrastruktur oder das Abstürzen des Datenerfassungsskriptes. In den Datensätzen befindet sich die Angabe der Zykluszeit, die beschreibt, wie lange die Maschine für die Produktion eines Teiles benötigt. Im Mittel beträgt diese für Maschine M69 22,83 Sekunden. Diese Zeit kann als Indikator für fehlende Datensätze gewählt werden. Es wurde ein Schwellenwert von 60 Sekunden zwischen zwei Daten festgelegt. Liegt die Zeitdifferenz darüber, wurde der jeweilige Datensatz für die Energieanalyse ignoriert. Daraus ergab sich bereits ein schlüssigeres Bild. Zudem existieren

Energiedifferenzen von 0 Wh, die ebenfalls aus Plausibilitätsgründen von der Analyse ausgeschlossen wurden.

Zusätzlich ist man nur an jenen Produktionsdurchläufen interessiert, bei denen ein ordnungsgemäßes Kunststoffteil produziert wurde. Datensätze mit Qualität = 3 werden also nicht berücksichtigt. Die Datenerfassung der Maschinen war an einigen Stellen unterbrochen. Für die Überprüfung musste eine Zeitdifferenz zwischen zwei Datensätzen gebildet werden. Bei der Arbeit mit Datenbanken sollte die Datenintegrität zu allen Zeitpunkten erhalten bleiben. Das bedeutet, dass die vorhandene Tabelle Produktionsprotokoll der Maschinen nicht mit der Zeitdifferenz ergänzt, sondern die separate Tabelle Energiedifferenz erstellt wurde, mit einem Verweis auf die ID im Produktionsprotokoll. Das hat den Vorteil, dass bei Fehlern die ursprünglichen Daten in Produktionsprotokoll unverändert bleiben und Änderungen durch Zurücksetzen der Energiedifferenz rückgängig gemacht werden können.

Energieverbrauch über die Zeit (M64)



Energieverbrauch über die Zeit (M67)



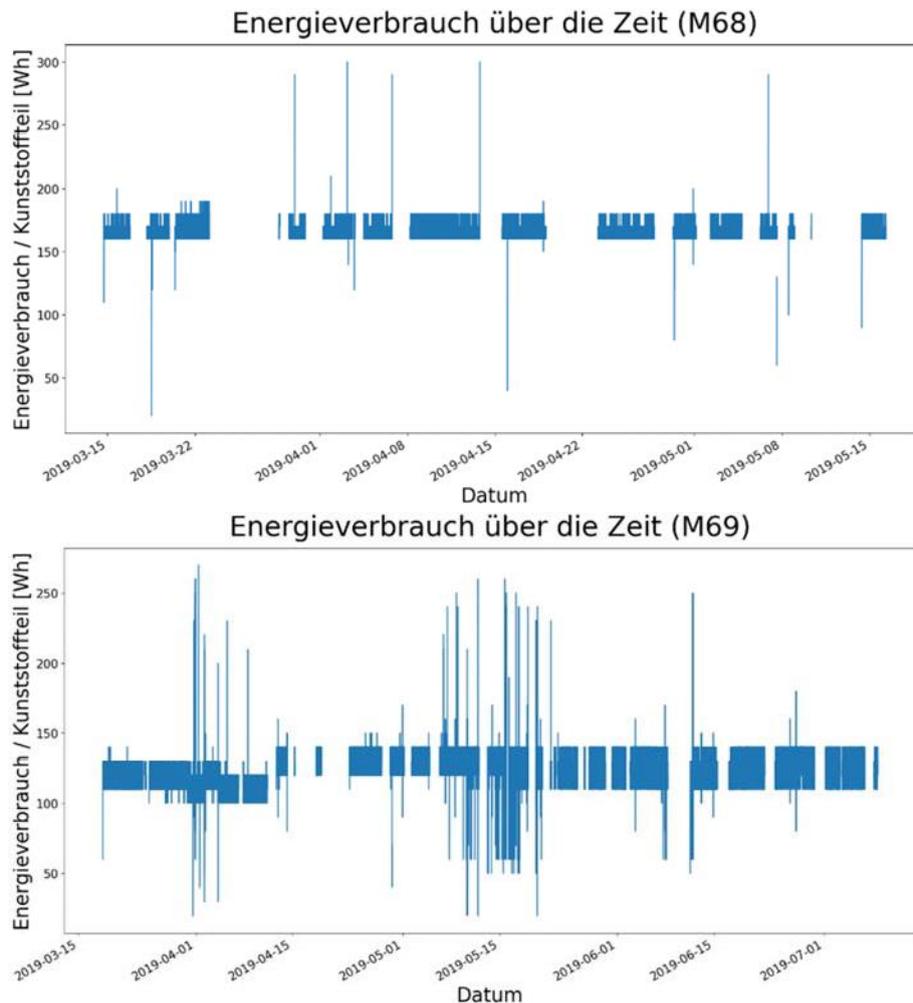


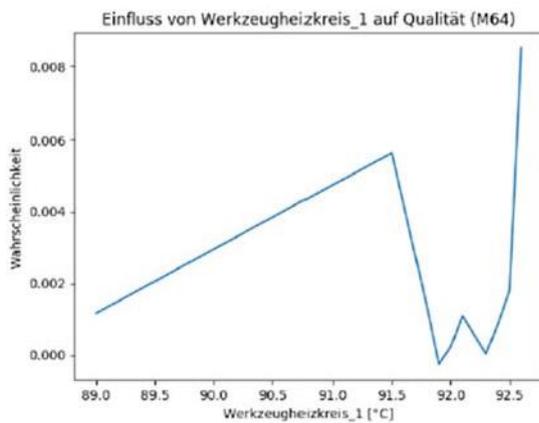
Abbildung 10 Energieverbräuche der Maschinen über die Zeit

Nach dem Plotten der Energieverbräuche fiel auf, dass der Energieverlauf weitgehend konstant ist. Das Rauschen im Signal lässt sich auf die Genauigkeit des Sensors von 10 Wh zurückführen. Das hat zur Folge, dass die Werte zwischen zwei bis drei 10er Schritten alternieren, z.B. bei M64 {150, 160, 170}. Es waren einige Spitzen zu erkennen, diese sind aber vermutlich auf fehlende Messdaten zurückzuführen, die die Filterung aus dem vorherigen Abschnitt nicht erfassen konnte. Bei Maschine M64 erhöht sich am 17.05.2019, um 13:12, schlagartig der Energieverbrauch von etwa 170 Wh auf 250 Wh. Für die Analyse mit Machine Learning ist dieses Gerät durch diesen signifikanten Unterschied am vielversprechendsten und kann als Validierer des späteren Algorithmus verwendet werden. In Rücksprache mit der Novapax GmbH konnten keine Änderungen in der Produktion zu diesem Zeitpunkt gefunden werden. Bei den Maschinen M67 und M69 existieren in kurzen Abschnitten vereinzelte Änderungen des Energieverbrauchs. Es bleibt zu analysieren, ob diese mit Änderungen in Features einhergehen.

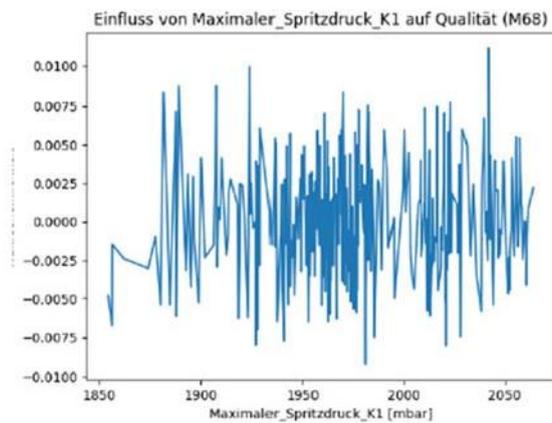
2.3 Ergebnisse der technischen Umsetzung und Handlungsempfehlungen

2.3.1 Interpretation der Diagramme

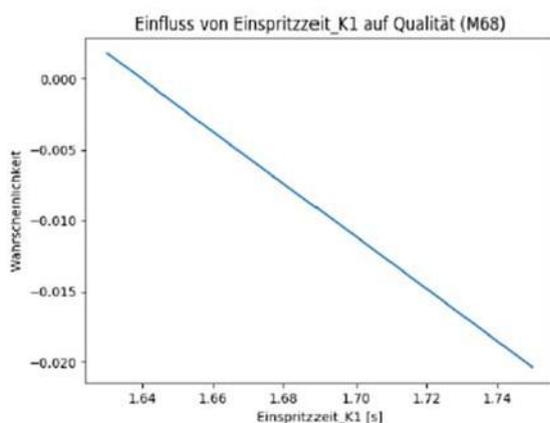
Beim Betrachten der Diagramme lassen sich verschiedene Arten von Kurven gruppieren. Bei der Verarbeitung mit LIME wurde bewusst eine vorzeitige Filterung der wichtigsten Features unterbunden, um erst bei der Interpretation entscheiden zu müssen, wie relevant ein Merkmal ist. Das führt dazu, dass im Nachgang eine geeignete Auswahl getroffen werden muss. Daher ist es wichtig, zuerst die Skala der y-Achse (Wahrscheinlichkeit oder Energie) zu betrachten. Als Voraussetzung für den jeweiligen Einbezug in die Auswertung wurde eine Relevanz von mindestens 1% für die Qualität und 2 Wh für die Energie gewählt.



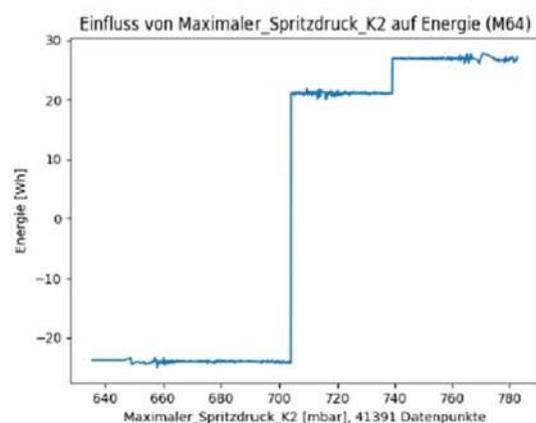
(a) Zu geringe Wahrscheinlichkeit



(b) Kein Muster erkennbar



(c) Nicht genügend Punkte



(d) Optimalfall

Abbildung 11 Auswertung der Kurven

Abbildung 11a zeigt beispielsweise einen klaren Anstieg in der Qualität zwischen 91.8°C und 92.4°C. Der Unterschied in der Wahrscheinlichkeit (0.005) ist jedoch zu gering, um eine fundierte Aussage zu machen. Ähnlich verhält sich Abbildung 11b. Dort kommt aber noch hinzu, dass sich von der Kurvenform her keine Präferenzen ausmachen lassen. Hier liegt ein klassisches Rauschsignal vor. Auch mit Filterung durch die minimale Anzahl an Datensätzen lässt sich keine Verbesserung erzeugen. Bis auf wenige Ausnahmen ist in diesem Diagramm auch nur ein Datensatz pro Punkt eingegangen. Vereinzelt existieren Diagramme wie Abbildung 11c, in denen sehr wenige X-Werte existieren. Hier ist bereits in den Rohdaten eine geringe Varianz festzustellen. Mögliche Ursachen hierfür könnten eine geringe Auflösung bei der Datenerhebung oder ein defekter Sensor sein. Den Optimalfall zeigt Abbildung 11d. Hier ist ein relativer Unterschied von 50 Wh vorhanden, die Kurven zeigen einen klaren Trend für optimalen Spritzdruck über 740 mbar und das Rauschen der Werte ist sehr gering. Eine weitere Überlegung sollte man für die Bewertung der Wichtigkeit in Betracht ziehen: Bei einigen Diagrammen befinden sich die Werte dauerhaft unter null und tragen demnach immer zu einer Schlechtteil-Klassifizierung bei. Daraus könnte man schlussfolgern, dass das Feature irrelevant ist, da bei keinem Wert ein Beitrag zur Gutteilerkennung geleistet wird. Entscheidend ist jedoch immer der Relativwert zweier Referenzen. Vielleicht ist es genau diese Differenz, die den Unterschied in der Klassifizierung macht.

2.3.2 Handlungsempfehlungen Maschine 64

Für Maschine M64 wurde mit ca. 168.000 Einträgen eine große Anzahl an Daten erhoben. Das hohe Verhältnis von 1,6% Schlechtteilen ist für die Auswertung der Qualität vorteilhaft. Dadurch sind detaillierte Ergebnisse zu erwarten. Ähnlich verhält es sich bei der Betrachtung der Energie. Im Gegensatz zu anderen Maschinen ist im Verlauf der Zeit ein plötzlicher Anstieg zu erkennen, durch den mit einer guten Klassifizierung zu rechnen war.

Merkmal	Empfehlung	Relevanzunterschied
Dosierzeit_K1 [s]	$x > 5.75$	0,02
Dosierzeit_K2 [s]	$x < 6.3$	0,14
Einspritzzeit_K2 [s]	$0,34 < x < 0,35$	0,29
Maximaler_Spritzdruck_K2 [mbar]	$685,0 < x$	0,04
Umschaltspritzdruck_K2 [mbar]	$685,0 < x$	0,06
Umschaltvolumen_K2 [cm ³]	$3,8 < x < 4,1$	0,04
Zykluszeit [s]	$x > 31,42$	0,15
Zylinderheizzone_1_K1 [°C]	$x = 190,0$ oder $x = 204,0$	0,14
Zylinderheizzone_1_K2 [°C]	$219,5 < x < 220,3$ oder $x = 224,5$	0,39
Zylinderheizzone_2_K2 [°C]	$x = 204,5$	0,05

Tabelle 6 Empfehlungen zur Optimierung der Qualität für Maschine M64

Tatsächlich hat Maschine M64 die vielversprechendsten Ergebnisse. Besonders klare Maxima lassen sich bei der Qualitätsanalyse (Tabelle 5.1) für die Merkmale Einspritzzeit_K2 und Zylinderheizzone_1_K2 erkennen. Mit 29% und 39% ist hier eine erhebliche Verringerung der fehlproduzierten Teile voraussagbar. Es wurden eindeutige Muster bei der Energie erkannt. Die höchste Optimierung wird für das Merkmal _Maximaler_Spritzdruck_K2_ prognostiziert: Hier ist eine Energiedifferenz von 45 Wh im Diagramm erkennbar. Nicht weniger herausragend ist die mögliche Einsparung von 25 Wh bei der Einspritzzeit_K2 und 15,5 Wh beim Umschaltspritzdruck_K2. Unter der Annahme, dass die Werte sich gegenseitig beeinflussen, ist eine Gesamtabstimmung mit 60Wh/Teil eine sehr konservative Prognose, die in der Praxis deutlich höher sein kann.

Merkmal	Empfehlung	Energiedifferenz [Wh]
Dosierzeit_K1 [s]	$x < 5,7$	5
Einspritzzeit_K2 [s]	$x > 0,37$	25
Massepolster_K2 [cm ³]	$x > 3,0$	10
Maximaler_Spritzdruck_K2 [mbar]	$x > 742$	45
Umschaltspritzdruck_K2 [mbar]	$x > 742$	15,5
Umschaltvolumen_K2 [mbar]	$x > 3,45$	13

Tabelle 7 Empfehlungen zur Optimierung der Energie für Maschine M64

2.3.3 Handlungsempfehlungen Maschine 67

Maschine M67 hat mit ca. 267.000 Einträgen den zweitgrößten Datensatz der betrachteten Maschinen. Im Betrachtungszeitraum wurde bereits die niedrige Fehlproduktionsquote von 0,4% erzielt. Mit 1.070 Schlechtteilen sind dennoch bei der Analyse gute Ergebnisse möglich. Der Energieverbrauch über die Zeit zeigt eine konstante Verteilung mit einzelnen Perioden von verringerter Energie. Daher war damit zu rechnen, dass Muster für eine Optimierung erkannt werden. Ein eindeutiges Ergebnis, wie bei Maschine M64, ist jedoch nicht zu erwarten. Die Auswertung der Qualität zeigt mehrere Verbesserungsmöglichkeiten, die zwischen 8% und 12% liegen. Dabei sind besonders die Zylinderheizzone präsent. Der Werkzeugheizkreis_1 zeigt ein klares Maxima bei 79,5 °C, an dessen Punkt die Wahrscheinlichkeit 9% höher ist als beim Minimum.

Merkmal	Empfehlung	Relevanzunterschied Qualität
Dosierzeit_K1 [s]	$x = 5,15$	0,03
Massepolster_K1 [cm ³]	$x = 8,4$	0,04
Umschaltspritzdruck_K1 [mbar]	$x > 712,0$	0,02
Werkzeugheizkreis_1 [°C]	$x = 79,5$	0,09
Zylinderheizzone_1_K1 [°C]	$267,1 < x < 269,3$	0,08
Zylinderheizzone_2_K1 [°C]	$x > 274,2$	0,08
Zylinderheizzone_4_K1 [°C]	$x > 279,2$	0,02
Zylinderheizzone_8_K1 [°C]	$x = 285,0$	0,12

Tabelle 8 Empfehlungen zur Optimierung der Qualität für Maschine M67

Bei der Energieoptimierung gibt es nur beim Umschaltvolumen_K1 mit 10Wh einen markanten Unterschied. Dennoch lässt sich aus den Ergebnissen aller Merkmale eine Einsparung von 20Wh/Teil prognostizieren.

Merkmal	Empfehlung	Energiedifferenz [Wh]
Integral (Druck über Zeit) [mbar/s]	$x > 870$	3
Massepolster_K1 [°C]	$x > 11,0$	5
Umschaltvolumen_K1 [cm ³]	$x < 22,0$	10
Zylinderheizzone_2_K1 [°C]	$x < 275,0$	2

Tabelle 9 Empfehlungen zur Optimierung der Energie für Maschine M67

2.3.4 Handlungsempfehlungen Maschine 68

Maschine M68 hat mit ca. 86.000 Datensätze die geringste Informationsmenge. Hinzu kommt, dass weniger als 1.000 Daten als Schlechtteil klassifiziert wurden. Beim Energieverbrauch im Verlauf der Zeit ist die Streuung der Werte bei dieser Maschine minimal und somit sind die Erwartungen bezüglich der Ergebnisse gering. Bei der Energie wurden lediglich Zusammenhänge mit der Zykluszeit

erkannt und diese ist mit 2,5 Wh auf einem geringen Level. Für den Energieverbrauch wird daher keine Einsparung durch veränderte Einstellungen prognostiziert. Anders sieht es bei der Qualität aus. Dort ist ein klarer Peak beim Werkzeugheizkreis_2 erkennbar, der einen Unterschied von 20% für die Qualitätsprognose macht. Mit jeweils 8% ist mit der Veränderung der Einspritzzeit_K2 und der Zylinderheizzone_1_K1 eine Verringerung der Fehlteile vorhersagbar. Insgesamt ist bei Maschine M68 im Praxistest nur wenig Optimierungspotential zu erwarten. Die wenigen Daten haben ihren Ursprung in der niedrigeren Produktion im Vergleich zu anderen Maschinen. Somit ist dieses Gerät ohnehin für die Gesamteinsparung weniger relevant.

Merkmal	Empfehlung	Relevanzunterschied Qualität
Einspritzzeit_K2 [s]	$x = 0,342$	0,08
Werkzeugheizkreis_2 [°C]	$x = 72,1$	0,20
Zylinderheizzone_1_K1 [°C]	$x = 194,5$	0,08

Tabelle 10 Handlungsempfehlungen zur Optimierung der Qualität für Maschine M68

Merkmal	Empfehlung	Energiedifferenz [Wh]
Zykluszeit [s]	$x > 33,5$	2,5

Tabelle 11 Handlungsempfehlung zur Optimierung der Energie für Maschine M68

2.3.5 Handlungsempfehlungen Maschine 69

Für Maschine M69 liegen mit ca. 275.000 die meisten Einträge in der Datenbank vor. Die hohe Fehlteilrate von 1,6% sorgt für eine gute Datengrundlage, um die Muster hinter guter und schlechter Qualität zu analysieren. Der Energieverbrauch im Verlauf der Zeit zeigt eine konstante Gerade, die jedoch abschnittsweise einen verringerten Energieverbrauch darbietet. In den Ergebnissen zur Qualität nehmen wie bei Maschine M67 die Zylinderheizzonen mit Werten zwischen 8% und 18% einen deutlichen Einfluss. Das Massepolster_K1 zeigt im Diagramm einen großen Rauschanteil. Im empfohlenen Bereich zwischen $6,1 \text{ cm}^3$ und $6,6 \text{ cm}^3$ ist jedoch eine klare Gerade sowie eine verbesserte Wahrscheinlichkeit für Gutteile zu erkennen. Beim Energieverbrauch existieren zwei Merkmale, die die Energie beeinflussen: Der Werkzeugheizkreis_3 und die Zykluszeit. In beiden Fällen gibt es jeweils einen Punkt, an dem die Energie rapide absinkt. Passt man die Maschineneinstellungen entsprechend an, sind Energieeinsparung von 10Wh/Teil zu erwarten.

Merkmal	Empfehlung	Relevanzunterschied Qualität
Dosierzeit_K1 [s]	$x = 5,26$	0,05
Massepolster_K1 [cm ³]	$6,1 < x < 6,6$	0,15
Temperiergerät_1_Werkzeug [°C]	$x = 25,8$	0,03
Zylinderheizzone_1_K1 [°C]	$x > 243,3$	0,18
Zylinderheizzone_2_K1 [°C]	$x = 244,1$	0,16
Zylinderheizzone_3_K1 [°C]	$x = 245,6$	0,08

Tabelle 12 Handlungsempfehlungen zur Optimierung der Qualität für Maschine M69

Merkmal	Empfehlung	Energiedifferenz [Wh]
Werkzeugheizkreis_3 [°C]	$x > 210$	8
Zykluszeit [s]	$x > 22,8$	9

Tabelle 13 Handlungsempfehlungen zur Optimierung der Energie für Maschine M69

2.3.6 Allgemeine Handlungsempfehlungen

Bei der Qualität fällt auf, dass häufig Merkmale entscheidend sind, die die Temperatur steuern. Insbesondere die Zylinderheizzone hat bei jeder Maschine ausnahmslos einen Einfluss auf die Qualität. Beim Blick auf die Herkunft der Daten, ist ein Zusammenhang schlüssig: Hat das Granulat für den Kunststoff nicht die richtige Temperatur, trocknet es unter Umständen zu schnell oder zu langsam. Fällt das produzierte Teil aus der Spritzgussform in die Ausgabe, können so im Nachhinein bei zu großer Hitze Verformungen eintreten, die das Produkt beschädigen. Bei zu niedriger Temperatur könnte das verflüssigte Granulat zu zäh sein und die Form nicht vollständig ausfüllen, was ebenfalls zur Aussortierung führt. Bei der Energie sind keine Gemeinsamkeiten unter den Maschinen erkennbar. Die Analyse muss also für jedes Gerät erneut erfolgen. Dennoch konnten signifikante Unterschiede in einigen Merkmalen herausgearbeitet werden. Die prognostizierte Gesamtersparnis für den Energieverbrauch wird in der folgenden Tabelle berechnet. Darauf aufbauend kann eine Vorhersage getroffen werden, wie viel CO₂ potentiell eingespart werden kann.

	Potentielle eingesparte Energie pro Teil [Wh]	produzierte Teile in Messzeitraum	Interpoliert auf ein Jahr	Energieeinsparung pro Jahr [kWh]
M64	60	168917	550488,438	33029,30625
M67	20	247052	805124,821	16102,49643
M68	0	86203	280929,42	0
M69	10	275286	897137,411	8971,374107

Tabelle 14 Vorhersage möglicher Energieeinsparungen pro Maschine

Basierend auf den Tabellen wurden für jede Maschine Annahmen über mögliche Einsparungen pro Teil getroffen. Der Messzeitraum betrug 112 Tage. Dieser Wert wurde auf 12 Monate interpoliert, um Aussagen über die Energieeinsparung pro Jahr zu machen. Insgesamt werden für die vier Maschinen eine Ersparnis von etwa 58.100 kWh prognostiziert. Auf Basis des Umweltbundesamtes wird für die CO₂-Reduktion ein Wert von 474g CO₂ pro kWh angenommen⁶. Daraus ergibt sich eine potentielle Einsparung von 27.540 kg CO₂ pro Jahr für die vier Maschinen.

2.3.7 Entwicklung eines künstlichen neuronalen Netzes (KNN)

Auf Basis der bisher durchgeführten Arbeiten wurde ein künstliches neuronales Netz (KNN) entwickelt, um die Leistung des Vorgängermodells zu verbessern. Dabei sollte untersucht werden, ob KNN ideale Maschineneinstellungen zur Energiereduzierung sowie der Verbesserung der Qualität finden können. Das KNN wurde ebenfalls auf die Maschinen M64 (168917 Datensätze), M67 (247052 Datensätze), M68 (86203 Datensätze) und M69 (275286 Datensätze) angewandt. Die Anzahl der Datensätze ist somit ausreichend, um ein KNN zu trainieren. Der Parameter „Schuss ID“ ist die ID des Datenpunkts und hat keine Relevanz für das Modell. „Maschine“ ist die Nummer der Maschine, daher wird sie zur Auswahl der relevanten Daten verwendet. Die Parameter „Integral“, „Programm“, „Werkzeug“, „Material“ und „Schuss nach Neustart“ haben keine Auswirkung auf das Modell, da sie durchgehend dieselben Werte haben und wurden entfernt. „Qualitaet“ sind die y-Werte für das Qualitätsmodell wobei „Energiezaehler“ und „Maschinen Zeitstempel“ die Basis für die Berechnung der y-Werte für das Energiemodell bilden. Die Qualitätsdaten bestehen meist aus guten Teilen. Die anderen Parameter können als x-Werte für beide Modelle verwendet werden. Alle Parameter mit einem Sternchen (*) sind Teil einer Zylindereinheit. Das Gerät besteht immer aus unterschiedlichen Wärmezonen, beginnend mit der Düse und endend an der Rohrleitung. Einige Parameter können an den Maschinen eingestellt werden, wie z. B. „Zylinderheizzone*”, „Umschaltvolumen*”, „Temperiergeraet Werkzeug*” und „Maximaler Spritzdruck*”. Daher bilden diese Parameter die unabhängigen Variablen ab. Andere Parameter wie „Zykluszeit“, „Massepolster*”, „Umschaltspritz-

⁶ Petra Icha, P., Kugs, G. Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990-2018. In: Umweltbundesamt – Climate Change 10 (2019), S. 29.

druck“, „Einspritzzeit“ und „Dosierzeit“ können lediglich gemessen werden und sind daher die abhängigen Variablen. Weiterhin waren Kameras installiert, um die Qualität der produzierten Teile anhand bestimmter Parameter zu messen und Unregelmäßigkeiten zu erkennen. Diese Daten wurden in einem Verzeichnis mit einer Reihe von Tabellendateien bereitgestellt. Um die Kameradaten mit den Daten in der Datenbank abzugleichen, wurden die Zeitstempel verglichen. Dies zeigte, dass nur 16.103 Datenpunkte übereinstimmende Zeitstempel hatten. Auf vier Maschinen verteilt, sind dies jedoch bei weitem nicht genug Daten, um ein neuronales Netzwerk zu trainieren. Darüber hinaus ist die Qualität der Kameradaten fraglich, da sie 6.262.794 Datensätze enthalten, mehr als sechsmal so viel wie die Datenbank. Aus diesem Grund sind die Kameradaten nicht brauchbar. Stattdessen wurde die in der Datenbank angegebene Qualitätsklassifizierung verwendet. Zunächst wurde eine einfache Dichte für das KNN für die jeweilige Energie- und Qualitätsmodellierung verwendet. Die Energiewerte werden in der Einheit Wh bereitgestellt und können einen beliebigen Wert annehmen. Typischerweise für diese Art von Ausgabewerten wird die Regression angewendet. Die Qualitätswerte sind jedoch binär, sodass sie entweder 0 (schlechte Qualität) oder 1 (gute Qualität) ergeben, daher wurde für dieses Modell eine binäre Klassifizierung verwendet. Das Modell hat eine erste verborgene Schicht von 32 Neuronen und eine zweite verborgene Schicht von 64 Neuronen. Die Komplexität des Netzwerks kann durch Ausprobieren verschiedener Konfigurationen gemessen werden. Ein zu vereinfachtes Netzwerk erzielt keine ausreichende Leistung.

Das KNN wurde nun unter Verwendung einer Aufteilung des Datensatzes trainiert. Wenn die Qualitätswerte 1 sind, werden Klassengewichte implementiert. Die Hyperparameter der Modelle wurden auf Überanpassung abgestimmt. Um eine bessere Verallgemeinerung zu erreichen, wurde das Netz reguliert. Die so trainierten Modelle bildeten dann die Basis für Parameter-Konfiguration-Optimierung. Im Folgenden wird das Klassifizierungsbeispiel für die Maschine M64 zur Veranschaulichung des Prozesses der Modellierung eines neuronalen Netzwerkmodells vorgestellt:

```
model = Sequential()
model.add(Dense(32, activation='relu', input_shape=(30,)))
model.add(Dense(64, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
opt = optimizers.SGD(lr=0.001)
model.compile(optimizer=opt,
              loss='binary_crossentropy',
              metrics=['accuracy'])

quality_model, quality_dataset = main.train_model(
    model=model,
    machine='M64',
    batch_sizes=(512,),
    epochs=1500,
)
```

Abbildung 1 Training des KNN

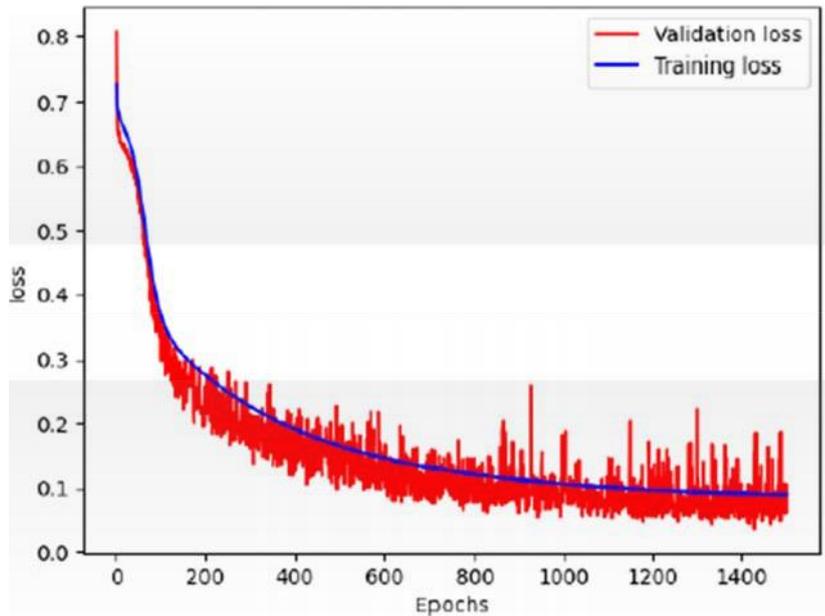
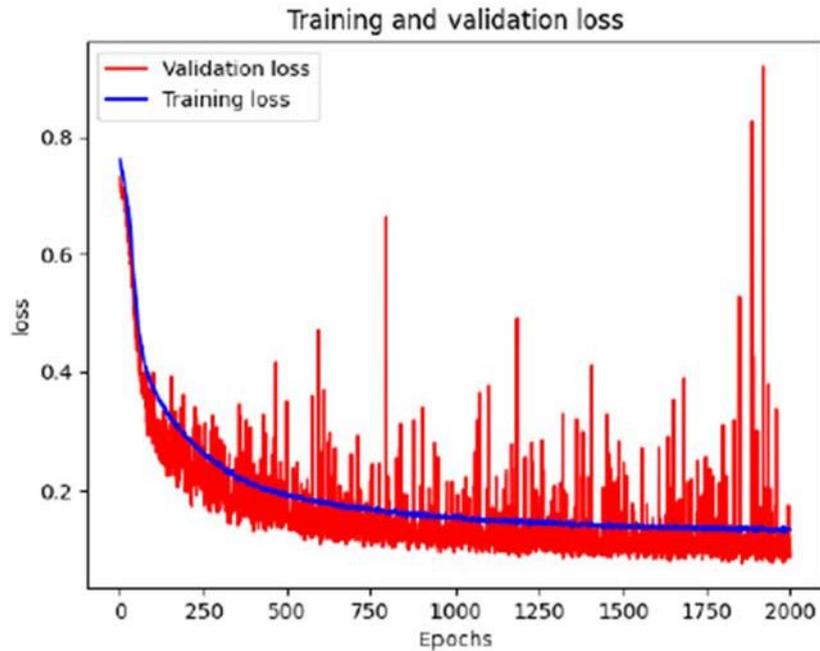


Abbildung 2 Modell der Qualitätsverluste der Maschine M64

Länge des y-Tests	16892
positiv	16617
negativ	275
wahr negativ	267
wahrhaft positiv	16450
falsch negativ	167 (0,010049948847565746)
falsch positiv	8 (0.02909090909090909)

Tabelle 1 Ergebnisse der Klassifizierung für Maschine M64

Die Modellausgabe zeigt, dass die Länge des Validierungssatzes 16892 Datenpunkte beträgt. Die Anzahl der positiven Datenpunkte, 16617, entspricht guten Teilen; 275 negative Daten repräsentieren die fehlerhaften Teile. Die Anzahl der echten Negative ist die Anzahl von korrekt erkannten fehlerhaften Teilen, während True Positives die Anzahl von korrekt erkannten guten Teile ist. Falsch negative sind die Teile, die gut sind, aber als fehlerhaft erkannt werden. Falsch positiv sind die fehlerhaften Teile, die als gut erkannt werden. Der prozentuale Anteil der falsch positiven Ergebnisse in diesem Modell beträgt 2,9 % und entspricht einer sehr guten Klassifizierung. Es ist jedoch wichtig, eine Regularisierung hinzuzufügen, um das Modell auf unsichtbare Daten zu verallgemeinern. Aus diesem Grund wird im nächsten Schritt die L2-Regularisierung hinzugefügt.



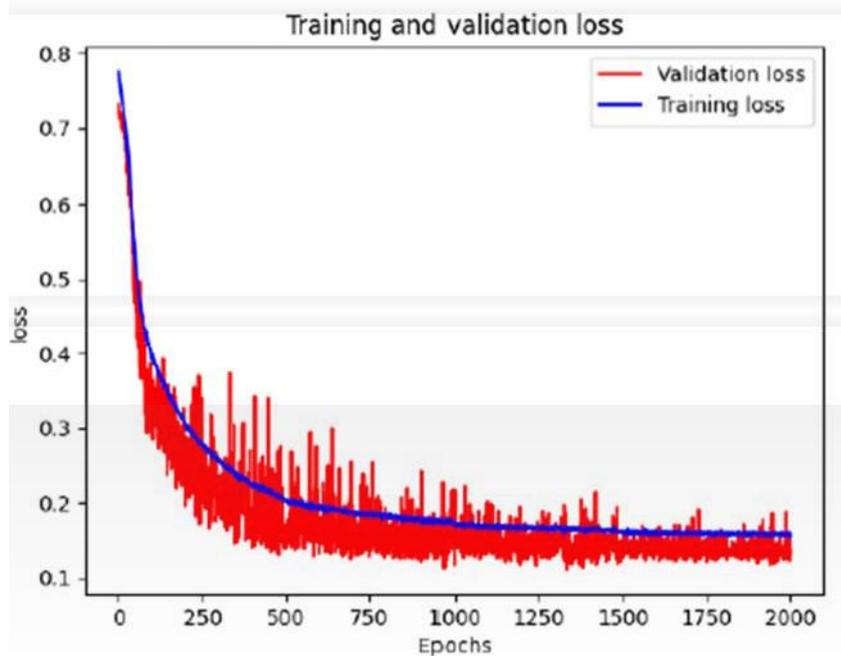
```
model = Sequential()
model.add(Dense(32, activation='relu', input_shape=(30,), kernel_regularizer=regularizers.l2(0.001)))
model.add(Dropout(0.1))
model.add(Dense(64, activation='relu', kernel_regularizer=regularizers.l2(0.001)))

model.add(Dropout(0.1))
model.add(Dense(1, activation='sigmoid'))
opt = optimizers.SGD(lr=0.001)
model.compile(optimizer=opt,
              loss='binary_crossentropy',
              metrics=['accuracy'])

quality_model, quality_dataset = main.train_model(
    model=model,
    machine='M64',
    batch_sizes=(256, 512, 1024, 2048),
    epochs=2000,
)
```

Abbildung 3 Trainings- und Validationsverlust

Wie in Abbildung zu sehen ist, zeigt das Modell mit mehr Iterationen eine höhere Varianz. Dies kann jedoch verhindert werden, indem die Chargengröße schrittweise erhöht wird. Bei diesem Modell wird die Chargengröße alle 500 Durchläufe verdoppelt.



```
model = Sequential()
model.add(Dense(32, activation='relu', input_shape=(30,), kernel_regularizer=regularizers.l2(0.001)))
model.add(Dropout(0.1))
model.add(Dense(64, activation='relu', kernel_regularizer=regularizers.l2(0.001)))
model.add(Dropout(0.1))
model.add(Dense(1, activation='sigmoid'))
opt = optimizers.SGD(lr=0.001)
model.compile(optimizer=opt,
loss='binary_crossentropy',
metrics=['accuracy'])

quality_model, quality_dataset = main.train_model(
model=model,
machine='M64',
batch_sizes=(256, 512, 1024, 2048),
epochs=2000,
class_weight_factor=4,
)
```

Abbildung 4 Trainings- und Validierungsverlust - Varianzreduktion

Wie in Abbildung 4 gezeigt, konnte die Varianz in den höheren Iterationen drastisch reduziert werden bei Erhöhung der Chargengröße. Die Rate der falsch positiven Ergebnisse ist mit 5,1 % deutlich höher als die Rate der falsch negativen mit 0,8 %. Da das Ziel ist, die Rate der falsch negativen Ergebnisse zu reduzieren. Ein Faktor 4 zur Änderung der Klassengewichte zeigte gute Ergebnisse für die Klassifizierungsraten mit einer Rate von falsch positiven 3,3 % und falsch negativen von 2,2 %.

Zur Reduzierung des Energieverbrauchs und der Maximierung der qualitativ hochwertigen Teile müssen zwei Modelle miteinander gekoppelt werden. Dazu wird im ersten Schritt die Vorhersagefunktion

$$\text{predict}(p_{\text{energy}}, p_{\text{quality}}) = \text{predict}(p_{\text{energy}}) * (1 - \text{predict}(p_{\text{quality}}))$$

verwendet, in welcher lokale Maxima gefunden werden sollen. Zum Auffinden lokaler Minima werden verschiedene Suchraumfunktionen angewendet. Anschließend wird jede Parameterkonfiguration mit dem Explanations-Tool Lime analysiert. Die Modellgenauigkeiten liegen in einem aussagekräftigen Bereich.

Machine	Number of Data Points	Quality Model	Energy Model
M64	16,892	97.70%	92.38%
M67	24,706	92.23%	88.73%
M68	8,621	98.70%	NaN
M69	27,529	84.50%	91.11%

Abbildung 5 Modellgenauigkeiten des Qualitäts- und Energiemodells

2.3.7.1 Handlungsempfehlungen M64

Der vielversprechendste Parameter ist Zylinderheizzone_5_K1, da mit leichter Abnahme der Qualitätsvorhersage der Energieverbrauch mit einem einzigen Schuss um 48,89 Wh reduziert werden kann. Dies entspricht einer jährlichen Energieeinsparung von 26.913 kWh. Zusätzlich reduziert Zylinderheizzone_2_K1 den Energieverbrauch nur um 6,29 Wh. Es gelingt jedoch auch, die Qualität um 4,87 % zu steigern. Eine Änderung wird empfohlen und würde eine jährliche Energieeinsparung von 3.463 kWh ergeben.

Parameter	Recommended Value	Quality Improvement in %	Energy Improvement in Wh
Zylinderheizzone_1_K1	206.504 °C	6.65	-4.27
Zylinderheizzone_2_K1	205.440 °C	4.87	6.29
Zylinderheizzone_5_K1	205.300 °C	-1.36	48.89
Umschaltvolumen_K2	4.429 cm ³	-5.69	32.28

Tabelle 2 M64: Qualitäts- und Energieempfehlung und prognostizierte Verbesserung

2.3.7.2 Handlungsempfehlungen M67

Als Parameteränderungen können nur Spritzdruck_K1 und Werkzeugheizkreis_1 für eine leichte Qualitätssteigerung sowie eine jährliche Energieeinsparung von 1,344 kWh empfohlen werden. Massepolster und Einspritzzeit können nur an den Maschinen gemessen jedoch nicht eingestellt werden.

Parameter	Recommended Value	Quality Improvement in %	Energy Improvement in Wh
Massepolster_K1	17.052 cm ³	0.13	7.14
Maximaler_Spritzdruck_K1	253.559 bar	0.14	0.43
Einspritzzeit_K1	1.503 s	0.05	0.08
Werkzeugheizkreis_1	80.000 °C	0.16	1.24

Tabelle 3 M67: Qualitäts- und Energieempfehlung und prognostizierte Verbesserungen

2.3.7.3 Handlungsempfehlungen M68

Mit nur 8621 Datensätzen war es nicht möglich, ein Energieregressionsmodell zu trainieren. Daher ist es nicht möglich, für diese Maschine auf Basis eines KNN eine Empfehlung abzugeben.

2.3.7.4 Handlungsempfehlungen M69

Für Maschine 69 konnten nur 2 Parameter gefunden werden, die sich verbessern ließen. Insgesamt könne jährlich ca. 520 kWh Energie eingespart werden.

Parameter	Recommended Value	Quality Improvement in %	Energy Improvement in Wh
Zylinderheizzone_1_K1	245.338 °C	0.06	0.11
Zylinderheizzone_4_K1	244.877 °C	0.04	0.47

Tabelle 4 M69: Qualitäts- und Energieempfehlung und prognostizierte Verbesserungen

3 Mögliche Anschlussarbeiten

Im nächsten Schritt gilt es, die erarbeiteten Ergebnisse an den Maschinen zu testen und deren Auswirkungen im Praxiseinsatz zu verifizieren. Diese Vorgehensweise ist elementar wichtig. Die erhobenen Daten bilden lediglich einen Ausschnitt der realen Umgebung, sowohl in der räumlichen als auch zeitlichen Komponente. Eine Korrelation zwischen zwei Parametern gibt nur einen Hinweis, kann aber keine Grundlage für einen Beweis zwischen Ursache und Wirkung geben. Während der Projektlaufzeit wurden weitere Maschinen in eine zentrale Datenerfassung aufgenommen. Die Daten in dieser Thesis basieren auf einem wenig zuverlässigen Skript, das durch einige Ausfälle eine lückenlose Erfassung behindert hat. Durch die Zentralisierungsmaßnahme beim Kooperationspartner Novapax können die Einflüsse in Tests der erarbeiteten Maschinenparameter optimal analysiert werden. Für weitergehende Forschung steht demnach ein deutlich größerer und qualitativ besserer Datensatz bereit, der detailliertere Ergebnisse liefern kann. Eine interessante Untersuchung für zukünftige Arbeiten ist ein globaler Ansatz, bei dem nicht die einzelnen Maschinen betrachtet werden, sondern deren Gesamtheit. So könnten etwa äußere Einflüsse wie eine Temperaturschwankung in der Produktionshalle erkannt und neue Optimierungsmöglichkeiten gefunden werden.

Softwareseitig gibt es ebenfalls Ansätze, die noch verfolgt werden können. Wie in der Evaluation der Algorithmen ausführlich dargestellt, bietet die Anwendung eines neuronalen Netzes auch eine große Erfolgsgarantie und könnte zu weiteren Ergebnissen führen. In der Interpretation der Modelle ist zu evaluieren, ob mit anderen Methoden eine größere Aussagekraft erzielt werden kann. In der Darstellung dieser Arbeit werden die Parameter unabhängig voneinander betrachtet. Eine mehrdimensionale Betrachtung, die Einstellungskombinationen analysiert, hat das Potential, weitere Energieeinsparungen zu offenbaren.

4 Fortschreibung des Verwertungsplanes

Die **HTW Berlin** wird die Projektergebnisse in Forschung und Lehre besonders in den Studiengängen Umweltinformatik, Betriebliche Umweltinformatik, Ingenieurinformatik sowie Life Science Engineering transferieren. Weiterhin förderte das Projekt den wissenschaftlichen Nachwuchs, indem Bachelor- und Masterarbeiten in das Projekt einbezogen und durch die wissenschaftlichen Mitarbeiter und Prof. Wohlgemuth betreut wurden:

- Till Michael Bussmann. „Netzwerkbasierte Maschinendatenerfassung und Auswertung für das Umweltmanagement in einem KMU der Kunststoffverarbeitenden Industrie“ (Masterarbeit). 2018.
- Maximilian Bönisch. „Konzept zur Einführung von maschinellem Lernen im Kontext der betrieblichen Nachhaltigkeit in mittelständischen Unternehmen“ (Masterarbeit). 2019.
- Jonas Scholten. „Machine Learning als Werkzeug zur Optimierung des CO₂- und Plastikverbrauchs bei der Produktion von thermoplastischen Kunststoffen eines Automobilzulieferers mittels Big Data Analyse“ (Masterarbeit). 2020
- Liam Alexander Risch „Künstliche neuronale Netze zur Optimierung des CO₂- und Plastikverbrauchs bei der Produktion von thermoplastischen Kunststoffen am Beispiel eine KMU (Arbeitstitel)“ (laufende Bachelorarbeit), geplante Abgabe: Oktober 2020

Novapax wird weitere Maschinen vernetzen und in das Machine Learning-Konzept integrieren. Somit kann innerhalb der Produktion langfristig eine signifikante Einsparung von CO₂, Energie und Abfall erreicht werden.

Beide Projektpartner stehen anderen KMU beim Einsatz von Machine Learning-Methoden beratend zur Seite. Geplant ist ein digitales Netzwerk mit best practise-Empfehlungen. Durch Veröffentlichungen in verschiedenen Print-Medien, Onlineplattformen und auf Messen/Konferenzen werden die Projektergebnisse dem interessierten Fachpublikum zur Verfügung gestellt:

- Druck von Flyern zum Nova-Projekt
- Erstellung einer Website: <http://nova.f2.htw-berlin.de/>
- Vorstellung des Projektes Nova auf dem Stand der DBU auf der Hannovermesse 2019
- Vorstellung des Projektes auf der Veranstaltung „Berlin macht's effizient“ des Bundesministeriums für Wirtschaft und Energie (BMWi) am 09.05.2019

- Artikel über das Projekt zum Thema transparente Produktionsprozesse im Umweltmagazin Bd. 50 (2020) Br. 06-07
- Einreichung einer Veröffentlichung im „Journal of cleaner production“

Beide Projektpartner planen Folgeanträge, um die Forschungsarbeiten um künstliche neuronale Netze zu erweitern. Damit könnte eine noch genauere Aussage über Handlungsempfehlungen getroffen werden. Allerdings ist die technische Umsetzung dieser Thematik sehr zeitintensiv, sodass der Einsatz von Fördermitteln vor dem personellen und finanziellen Hintergrund notwendig ist.

5 Während des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt

Es liegen keinerlei Hinweise vor, die darauf hindeuten, dass die im Projekt fokussierten Themenfelder ihre Relevanz verloren haben bzw. Lösungen und Konzepte entwickelt wurden, die das Forschungsprojekt obsolet werden ließen.

6 Anhänge

Optimierung der Anlagen in Bezug auf Ressourcenverbrauch und Abfallaufkommen

Dazu werden die Produktionsprozesse in einem Stoffstrommanagementsystem erfasst, aufbereitet und analysiert, um Optimierungspotenziale der Stoff- und Energieströme aufzudecken und Verbesserungen in Bezug auf Material, Abfall und Energie umzusetzen.

Reduzierung von CO2- Emissionen

Ein besonderes Augenmerk liegt auf der Reduzierung der CO2-Emissionen. Dafür werden geeignete Schnittstellen für den Im- und Export der Maschinendaten entwickelt, um diese prozessübergreifend verarbeiten zu können. Neben den Betriebsdaten sind auch Umgebungsfaktoren (Temperaturschwankungen, Hindernisse etc.) zu berücksichtigen und in die Optimierungsanalyse einzubeziehen.

<http://nova.f2.htw-berlin.de>



Nova

Entwicklung einer neuen Methode zur Ressourceneinsparung durch prozessorientiertes, digitales Stoffstrommanagement in KMU



Hochschule für Technik und
Wirtschaft Berlin



**NOVAPAX Kunststofftechnik
Steiner GmbH & Co. KG**

NOVAPAX

Dafür ist maschinelles Lernen dann geeignet, wenn ausreichend Trainingsdaten zur Verfügung stehen. Aufgrund des hohen Materialdurchsatzes bei der Fertigung der Kunststoffteile werden die Voraussetzungen für die Entwicklung geeigneter Lernverfahren erfüllt. Dabei sollen die Daten nicht nur offline, sondern bereits im laufenden Betrieb trainiert werden. Daneben werden auch indirekte Faktoren (z.B. Wartungsintervalle, mechanische Störungen) berücksichtigt.

Intelligente Digitalisierung von Produktionsprozessen

Die verschiedenen Fertigungsprozesse, welche sich durch starke Nichtlinearitäten und zeitinvariantes Verhalten auszeichnen, sollen mit Machine Learning und selbstlernenden Steuerungen analysiert werden.

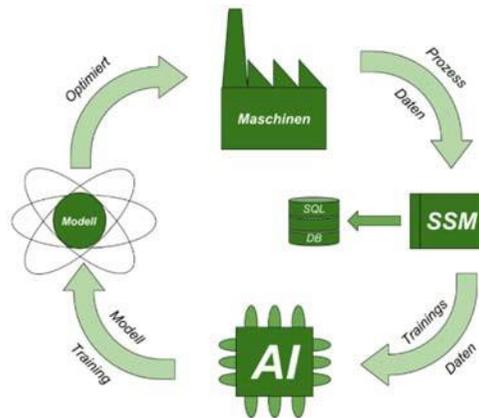


Abbildung 1 Schematische Darstellung des Projektlaufes (Bönisch, M., Edikh, V., 2018)

Projektbeschreibung

Um den Prozess der Abfall- und Ausschussminimierung sowie die Energieeinsparung optimieren zu können, soll im Forschungsprojekt Nova eine intelligente automatisierte Maschinendatenerfassung mit einem selbstlernenden Steuerungssystem auf Basis von künstlicher Intelligenz realisiert werden.

Die wissenschaftliche Herausforderung besteht darin, komplexe industrielle Prozesse mit modernen Methoden der Informatik abzubilden und durch Algorithmen des maschinellen Lernens zu optimieren, um erhebliche Ressourceneinsparungen für den gesamten Produktionsprozess zu erzielen.



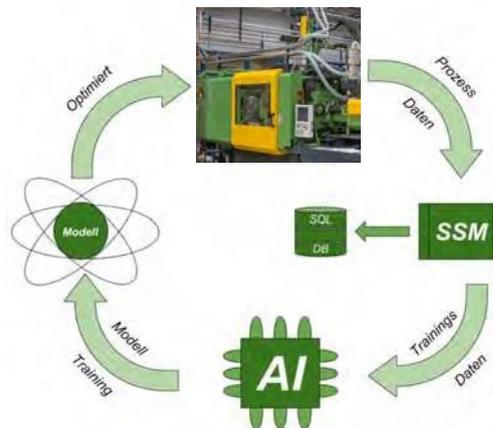
Umweltinformatik

Hochschule für Technik und Wirtschaft Berlin

htw.

Projekt NOVA

Entwicklung einer neuen Methode zur Ressourceneinsparung durch prozessorientiertes, digitales Stoffstrommanagement in KMU
Entwicklung einer neuen Methode zur Ressourceneinsparung durch prozessorientiertes, digitales Stoffstrommanagement in KMU



In Kooperation mit NOVAPAX Kunststofftechnik Steiner GmbH & Co. KG

htw.

Projekt NOVA

Projektbeschreibung:

Um den Prozess der Abfall- und Ausschussminimierung sowie der Energieeinsparung optimieren zu können, soll im Forschungsprojekt Nova eine intelligente automatisierte Maschinendatenerfassung mit einem selbstlernenden Steuerungssystem realisiert werden.



In Kooperation mit NOVAPAX Kunststofftechnik Steiner GmbH & Co. KG

htw.

Projekt NOVA

Zielstellung

Intelligente Digitalisierung von Produktionsprozessen

Optimierung der Anlagen in Bezug auf Ressourcenverbrauch und Abfallaufkommen



In Kooperation mit NOVAPAX Kunststofftechnik Steiner GmbH & Co. KG

htw.

Projekt NOVA

Anlagenoptimierung hinsichtlich Ressourcenverbrauch und Abfallaufkommen

- Abbildung und Analyse der Maschinendaten mittels einer Stoffstrommanagement-Software
- Besonderer Fokus liegt in der Minderung des CO₂-Ausstoßes.
- Zur Erfassung der notwendigen Maschinendaten werden geeignete Schnittstellen für den Datenimport entwickelt.



In Kooperation mit NOVAPAX Kunststofftechnik Steiner GmbH & Co. KG

htw.

Projekt NOVA

Intelligente Digitalisierung von Produktionsprozessen

- Analyse der Fertigungsprozesse, welche sich durch starke Nichtlinearitäten und zeitinvariantes Verhalten auszeichnen
- Modellbildung mit Methoden des maschinellen Lernens
- Kombination dieser Modelle mit dem Stoffstrommanagementsystem



In Kooperation mit NOVAPAX Kunststofftechnik Steiner GmbH & Co. KG

htw.

Projekt MFCA mobil

Mobile Unterstützung der Materialflusskostenrechnung in KMU

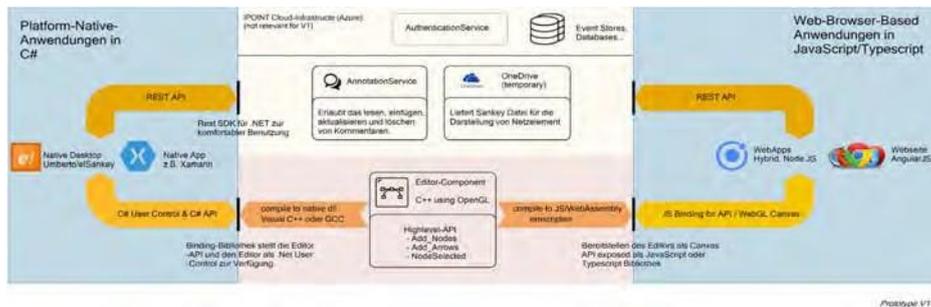


Abbildung 1: Schematische Darstellung der Projektarchitektur (ifu Hamburg GmbH, 2018)

In Kooperation mit:
ifu Institut für Umweltinformatik & Umweltkompetenzzentrum Rhein-Neckar e.V. (UKOM)

htw

Projekt MFCA mobil

Projektbeschreibung:

Im Projekt MFCA mobil soll die Weiterentwicklung eines bereits als Prototyp bestehenden Werkzeugs zur mobilen Erfassung und Visualisierung von Stoffströmen in Unternehmen zu einem attraktiven Softwaresystem für die Unterstützung von Ressourceneffizienz in Unternehmen realisiert werden. Das Projekt baut unter anderem auf dem DBU-Projekt Mobile Unterstützung eines prozessorientierten Stoffstrommanagements in KMU (MOPS) auf.



In Kooperation mit:
ifu Institut für Umweltinformatik & Umweltkompetenzzentrum Rhein-Neckar e.V. (UKOM)

htw

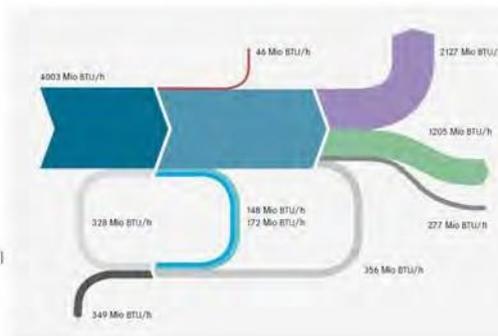
Projekt MFCA mobil

Zielstellung

Mobiler Ansatz des Stoffstrommanagements mit der Methode der Materialflusskostenrechnung (MFCA)

Beispiel: Oxyfuel- Verbrennungsverfahren

- Thermische Energie von Kohle [Mio BTU/h]
- Wärmeverlust Heizkessel [Mio BTU/h]
- Dampferzeugung [Mio BTU/h]
- Energiebedarf Oxyfuel-Anlage [Mio BTU/h]
- Kompressionskraft [Mio BTU/h]
- Generator Output [Mio BTU/h]
- Thermodynamische Verluste [Mio BTU/h]
- Wärmegehalt Abgase [Mio BTU/h]
- Latenter Wärmegehalt der Abgase [Mio BTU/h]
- Zurückgewonnene thermische Energie [Mio BTU/h]
- Kompressionsverlust durch Abgas [Mio BTU/h]



Das Diagramm basiert auf fiktiven Werten.
1000 BTU/h = 293 W

In Kooperation mit:
ifu Institut für Umweltinformatik & Umweltkompetenzzentrum Rhein-Neckar e.V. (UKOM)

htw.



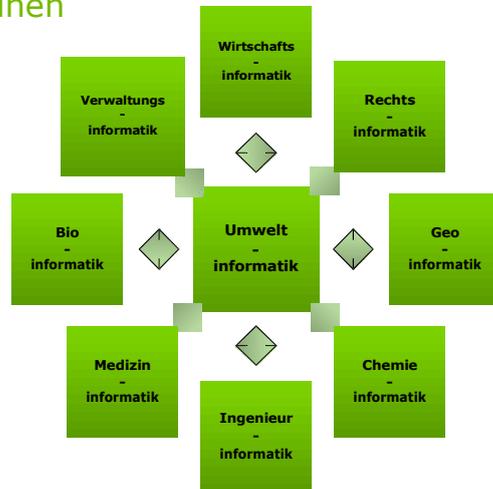
htw. Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

Umweltinformatik

Fachbereich
Ingenieurwissenschaften
Technik und Leben

Umweltinformatik - UI

Beziehung der Umweltinformatik zu anderen Teildisziplinen



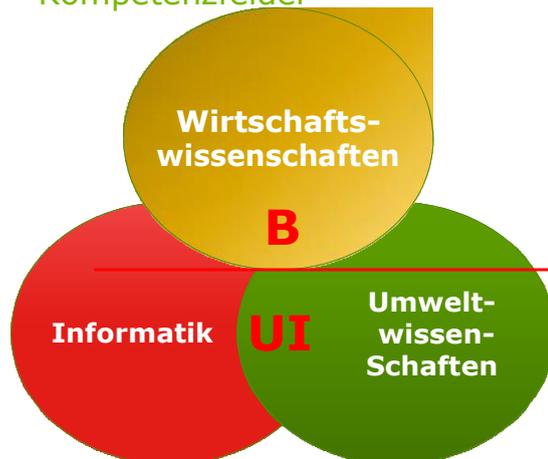
11

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist UI- Umweltinformatik/ BUI – Betriebliche Umweltinformatik?

Kompetenzfelder



- Einzigartige Mischung von Kompetenzen
Master BUI
(Schwerpunkt Informatik)
- Interdisziplinäres Umfeld
Bachelor UI
(Absolvent als Moderator)
- ▲ Hohes Gestaltungspotential
(Absolvent als Gestalter)

12

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.



Der Bachelor-Studiengang UI

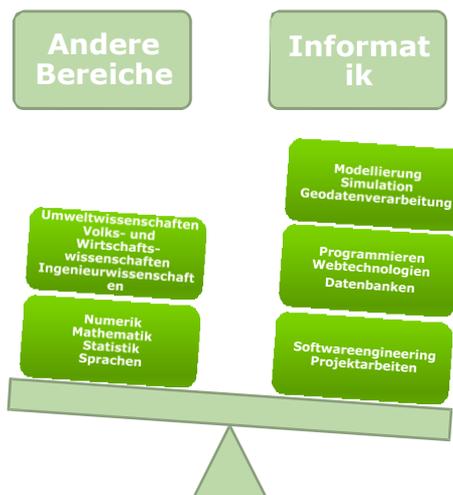
Umweltinformatik

13

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Umweltinformatik Gewichtung

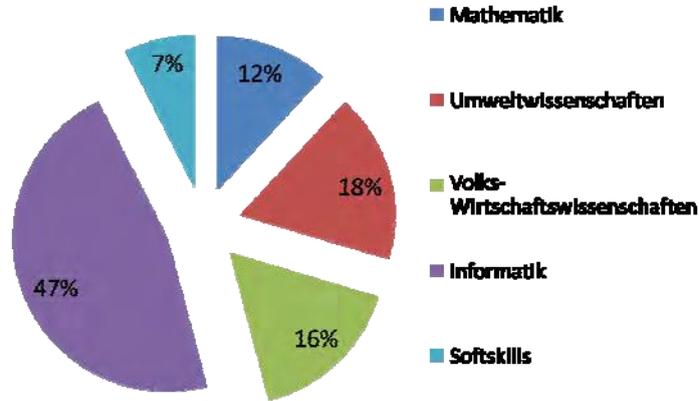


14

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Bachelorstudiengang UI Inhalte



15

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Bachelorstudiengang UI Inhalte



1. Semester	Mathematik 1 (Analysis)	Ökologie und Biologie	Grundlagen der Verwaltungs- und Wirtschaftswissenschaften 1	Einführung in die Informatik	Programmierung 1	1. Fremdsprache
2. Semester	Mathematik 2 (Lineare Algebra und Diskrete Strukturen)	Umweltchemie	Grundlagen der Verwaltungs- und Wirtschaftswissenschaften 2	Grundlagen der Kostenrechnung	Programmierung 2	Modellierung in der Informatik
3. Semester	Numerische Algorithmen	Umweltanalytik und -messverfahren	Umwelt- und Geoinformationssysteme 1	Datenbanksysteme	Webtechnologien	Projekt: Umwelt - Informatik - Gesellschaft
4. Semester	Deskriptive Statistik und Stochastik	Umweltphysik und Geographie	Umwelt- und Geoinformationssysteme 2	Simulation von Umweltsystemen	Software Engineering*	Verteilte System und Komponentenarchitektur AWE 1: wissenschaftliches Arbeiten
5. Semester	Ingenieurwissenschaftliche Grundlagen und CAD	Ausgewählte Kapitel der Umweltwissenschaften	Umweltrecht	Vertiefung Umweltinformatik	Software Ergonomie	Projekt: Projektmanagement (inkl. Software-projekt) AWE 2**
6. Semester	Praxisphase: Fachpraktikum	Bachelorarbeit / Kolloquium		Mathematikwissenschaften Verwaltungs- und Wirtschaftswissenschaften Umweltwissenschaften Ingenieurwissenschaften Informatik Allgemeinwissenschaftliches Ergänzungsfach Projekt Fachpraktikum und Schreiben der Bachelorarbeit		

16

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist Umweltinformatik?

Tätigkeiten – Visualisierung 3-D-
Geländedarstellung



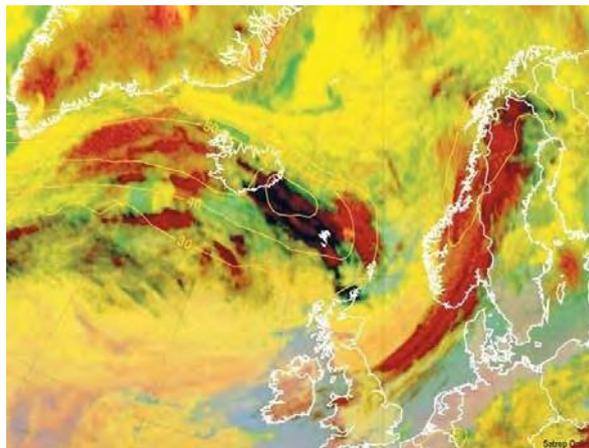
17

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist Umweltinformatik?

Tätigkeiten – Visualisierung 3-D-
Geländedarstellung



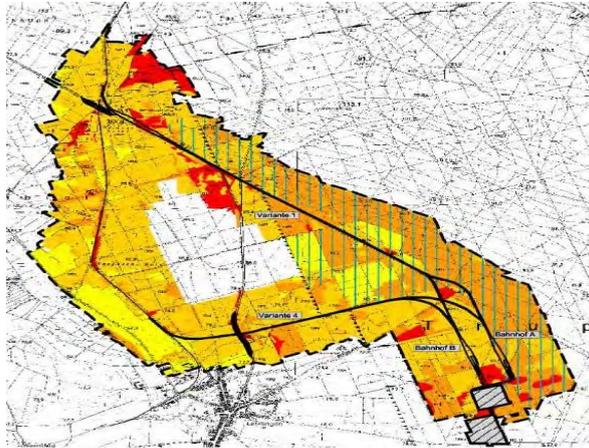
18

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist Umweltinformatik?

Tätigkeiten – GIS-Anwendung Umweltverträglichkeitsprüfung



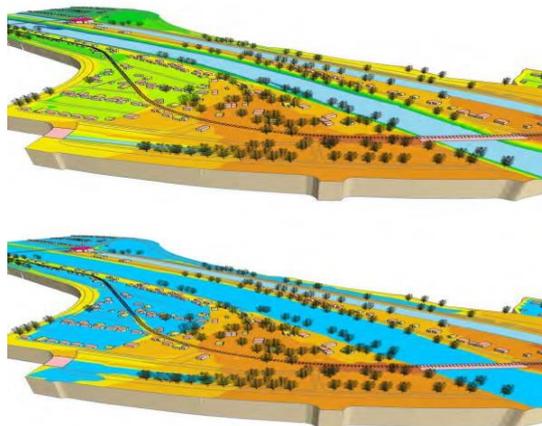
19

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist Umweltinformatik?

Tätigkeiten – Modellieren und Simulieren



20

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.



Der Master-Studiengang BUI

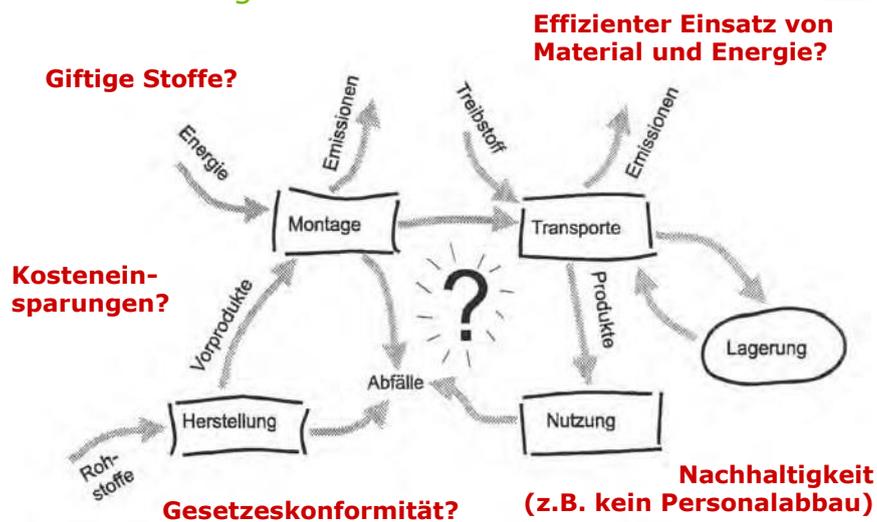
Betriebliche Umweltinformatik

21

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist BUI? Problemstellung



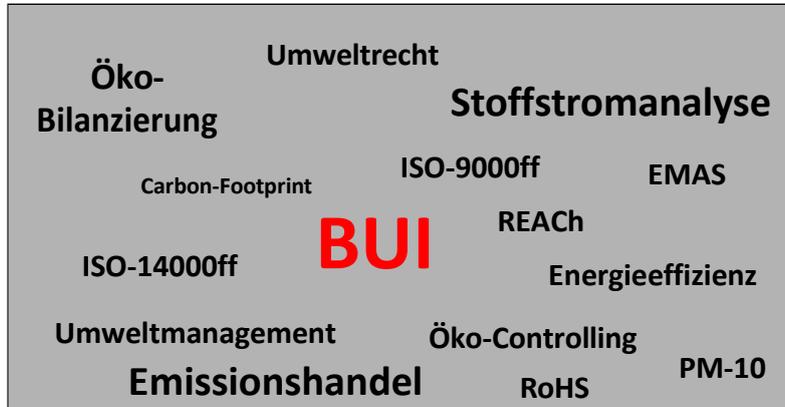
22

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist BUI?

Themen



➔ Nutzung/Schaffung von IT-Lösungen

23

Vorstellung Studiengang Umweltinformatik
Prof. Dr.-Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist BUI?

Aufgabenfelder



-  Harmonisierung von Ökonomie und Ökologie
-  Betrachtung von verschiedenen Aspekten der Nachhaltigkeit
-  Reduzierung der Komplexität betrieblicher Tätigkeiten
-  Sicherstellung der Legal Compliance
-  Bereitstellung zielgerichteter Informationen für operative, taktische und strategische Managemententscheidungen

➔ Unterstützung mit Mitteln der Informatik

24

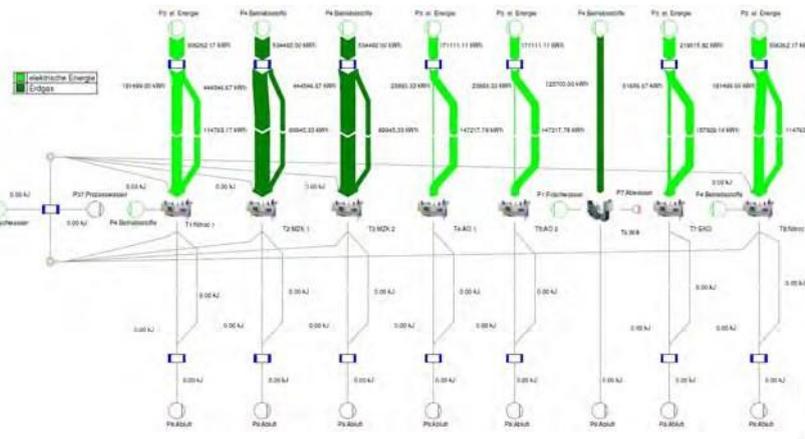
Vorstellung Studiengang Umweltinformatik
Prof. Dr.-Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist BUI? Projekte



Entwicklung und Anwendung von Software zur Energieoptimierung von Produktionsprozessen



27

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Was ist BUI? Projekte



Entwicklung einer Webapplikation zur Bewertung der Nachhaltigkeit von Biokraftstoffen

The screenshot shows a web application interface for evaluating the sustainability of bioenergy. The interface is divided into several sections. On the left, there is a navigation menu with options: 'Basic Data', 'Cultivation', 'Processing', 'Social criteria', 'Validation/Calculation', and 'Results'. The main content area is titled 'Land use' and contains input fields for 'Yield', 'Ecozone', 'Annual rainfall', and 'Winter type precipitation distribution?'. Below this, there is a 'Soil information' section with fields for 'USDA Soil Order' and 'Mean slope'. The interface also includes a 'Reset entered data' button and a 'Please lock up here' message.

28

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Sie erinnern sich?



Der Klimawandel kommt.
Wir machen mit!

Umweltschutz an der HTW -
von der Theorie zur Praxis -
Umweltmanagement jetzt!

29

Vorstellung Studiengang Umweltinformatik
Prof. Dr. -Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Masterstudiengang BUI

Masterstudium besteht aus 6 Lehrgebieten



1. Softwaretechnische Grundlagen:

- Webtechnologien und -programmierung
- Software- und Webarchitekturen
- Aktuelle Entwicklungstrends für BUIS



Grundlagen der Modellbildung und Simulation:

- Grundlagen der Systemtheorie und Modellbildung
- Prozesssimulation
- (WP-Modul Vertiefung Quantitative Methoden)



Technische, rechtliche und gesellschaftliche Management-Aspekte in der BUI:

- Methoden und Werkzeuge in BUI-Projekten
- Rechtliche und gesellschaftliche Aspekte in der BUI
- Betriebliche Umwelttechnik

Masterstudiengang BUI

Masterstudium besteht aus 6 Lehrgebieten



4. Umweltmanagement:

- Umweltorientierte Betriebswirtschaftslehre
- Betriebliches Umweltmanagement
- Betriebliche Umweltkostenrechnung



5. Betriebliche Anwendungen und BUIS:

- Informations- und Wissensmanagement
- Grundlagen betrieblicher Umweltinformationssysteme (BUIS)
- Stoffstrommanagement und LCA



6. Projekte

- Praxisprojekt
- Forschungsprojekt

Studiengänge UI/BUI

Kennzeichen



**Schwerpunkt: Informatik
(Angewandte und Praktische Informatik)**



**Orientierungswissen:
Ingenieurwissenschaften, Umweltwissenschaften und
Wirtschaftswissenschaften**



**Soft Skills:
Präsentations- und Kommunikationstechniken,
Moderation, Englisch**

**Breitgefächert
und interdisziplinär!**



Hoher Praxisanteil (Master- und Bachelor)



Studienangebote:

**Umweltinformatik (Bachelor),
Betriebliche Umweltinformatik (Master)**

**In dieser Form
einzigartig in
Europa!**

Studiengänge UI/BUI Projektpartner



33

Vorstellung Studiengang Umweltinformatik
Prof. Dr.-Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.

Studiengang BUI Weitere Informationen

 **Studiengangssprecher und Studienfachberater**
Prof. Dr.-Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de
030 5019 3308

 **www.ui-berlin.de**

34

Vorstellung Studiengang Umweltinformatik
Prof. Dr.-Ing. Jochen Wittmann
Jochen.Wittmann@htw-berlin.de

htw.



htw Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

**Vielen Dank für Ihre
Aufmerksamkeit!**

Energieeffizienz und Industrie 4.0 in KMU

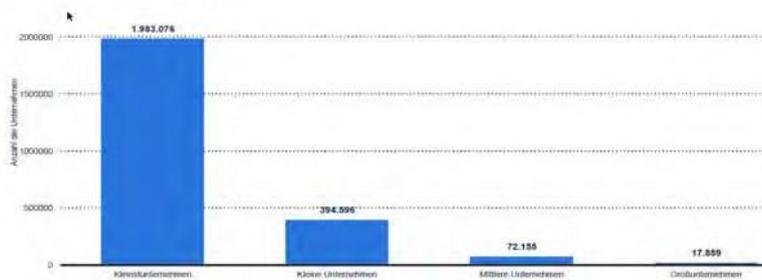
Carola von Holt
NOVAPAX Kunststofftechnik
Steiner GmbH & Co. KG

Martina Willenbacher
Hochschule für Technik und
Wirtschaft Berlin

Berlin, 09.Mai 2019

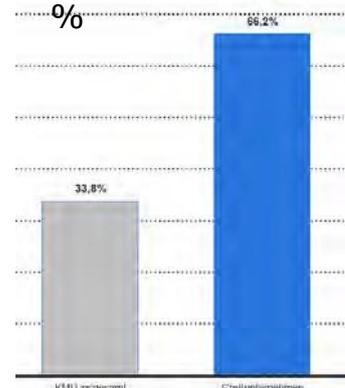
KMU in Deutschland

- Anteil: 99,5 % der deutschen Unternehmen (1,98 Mio.)



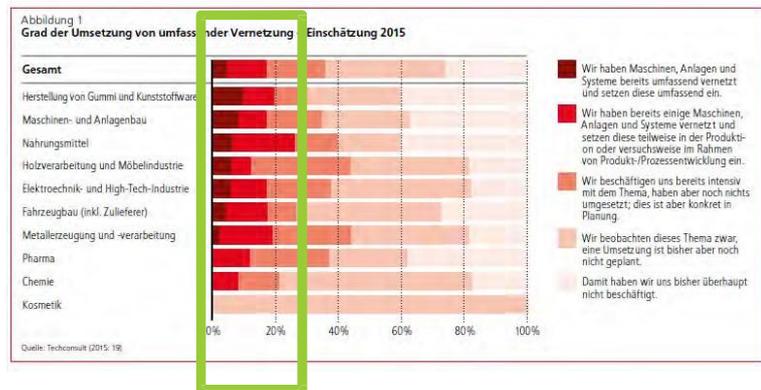
Statistisches Bundesamt, <https://de.statista.com/statistik/daten/studie/731859/umfrage/>

- Umsatz: 33,8 %



KMU - Digitalisierung

- Anteil **umfassend vernetzter** KMU: 5,5 %
- Anteil **teilweise vernetzter** KMU: 13 %



KMU – Hemmnisse der Digitalisierung

- Innovationshemmnisse: Risiko, Kosten, Know-how



Bundesministerium für Wirtschaft und Energie: <https://www.bmwi.de/Redaktion/DE/Infografiken/Mittelstand/infografik-mittelstand-digital.html>

Forschungsprojekt Nova

- Entwicklung einer neuen Methode zur Ressourceneffizienz durch prozessorientiertes, digitales Stoffstrommanagement in KMU
- Konsortium:  
Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences
- Mittelgeber: Deutsche Bundesstiftung Umwelt DBU
- Laufzeit: 01.04.2018 bis 30.09.2020



Nova - Hintergrund

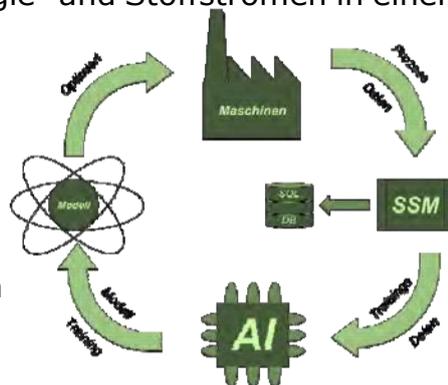
- Fehlende dezentrale, mobile Datenerfassung
- Implementierung, Pflege und Wartung eines kostenintensiven Umweltinformationssystems ist verbunden mit hohen Personalkosten.
- Machine Learning und Künstliche Intelligenz sind komplexe Themenbereiche, die ein hohes Maß an Fachwissen und Investitionen erfordern.

Nova - Konzept

- Implementierung eines Stoff- und Energieflussmanagementsystems zur Analyse, Bewertung und Visualisierung und Ableitung von Maßnahmen zur Optimierung von Energie, Material und Schadstoffen
- Analyse von Fertigungsprozessen, die durch starke Nichtlinearitäten und zeitinvariantes Verhalten bei großen Datenmengen gekennzeichnet sind
- Abbildung komplexer industrieller Prozesse mit modernen Methoden der Informatik und deren Optimierung durch Algorithmen des maschinellen Lernens, um erhebliche Ressourceneinsparungen für den gesamten Produktionsprozess in KMU zu erzielen

Nova – Konzept II

1. Automatisierte, intelligente Datenerfassung der Produktionsprozesse
2. Abbildung der Prozesse mit ihren Energie- und Stoffströmen in einem Umweltmanagementsystem
3. Datenaufbereitung und Erstellung von Trainings- und Testdatensätzen
4. Modelltraining und Validierung
5. Optimierung von Produktionsprozessen durch Modellvorhersagen



Nova - Ziele

- Optimierung von Anlagen und Maschinen im Hinblick auf Ressourcenverbrauch und Abfallerzeugung mit computergestütztem Stoffstrommanagement und Verfahren der künstlichen Intelligenz
- Implementierung einer nachhaltigen Digitalisierung von Fertigungsprozessen durch maschinelles Lernen zur Reduzierung schädlicher Umweltauswirkungen
- Abbildung und Analyse von Maschinendaten mit Hilfe von Materialflussmanagementsoftware mit besonderem Schwerpunkt auf die Reduzierung der CO2-Emissionen

Diskussion



NOVAPAX



Projekttreffen Nova

Projektlaufzeit: 01. März 2018 – 31. Dezember 2020

Projektträger: Deutsche Bundeshilfe Umwelt (DBU)

Förderkennzeichen: 34589/10

Berlin, 13. August 2020

Ausgangssituation und Projektziele

- Digitalisierung von Produktionsprozessen anhand von ausgewählten Maschinen (Maschinenvernetzung)
- Entwicklung von Machine Learning-Algorithmen zur Optimierung der Anlagen hinsichtlich Energie und Abfall
- Ableiten von ersten Handlungsempfehlungen auf Basis der theoretischen Ergebnisse

Vorgehen und Grundannahmen

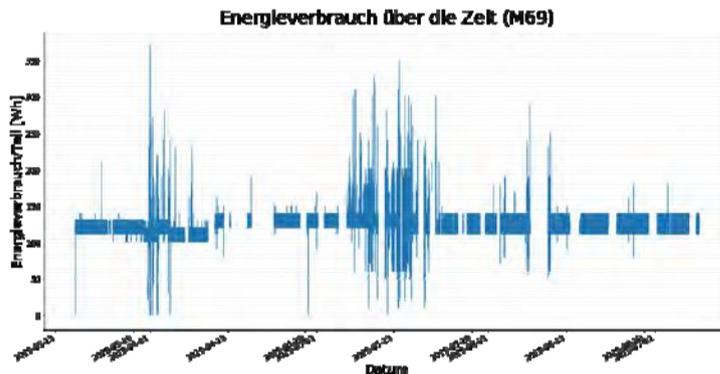
- Vorgehensweise:
 1. Analysieren und Validieren der Datensätze
 2. Mittels Machine Learning-Algorithmen Muster finden, die zu unerwünschtem Verhalten führen
 3. Handlungsempfehlung, welche Parameter verändert werden sollen
- Annahmen bei Machine Learning:
 1. Es gibt einen Zusammenhang zwischen Eingangs- und Ausgangswerten.
 2. Es sind genug Daten vorhanden, um den Zusammenhang zu erkennen.

Vorraussetzungen

- 4 Maschinen
- Ca. 800.000 Daten in einem Zeitraum von 112 Tagen
- Erhobene Daten:
 - Maschinen_Zeitstempel, Schuss_ID, Maschine, Werkzeug, Programm, Material, Qualität, Schuss_nach_Neustart, Zykluszeit, Massepolster, Umschaltvolumen, Umschaltspritzdruck, Maximaler_Spritzdruck, Einspritzzeit, Dosierzeit, Zylinderheizzone, Werkzeugheizkreis, Temperiergeraet_Werkzeug, Integral (kumulierter Druck über die Zeit), Energiezähler

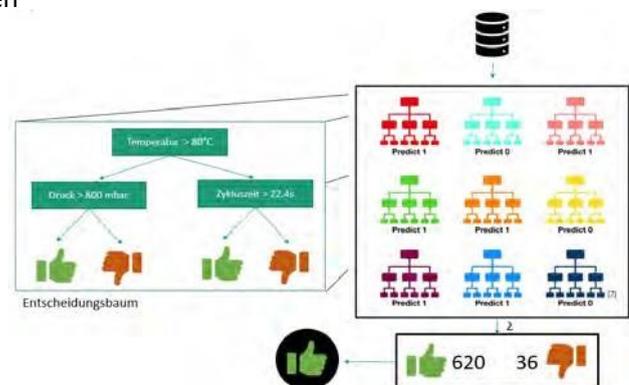
Datenvorverarbeitung

- Validierung der Daten auf Vollständigkeit
- Berechnung des Energieverbrauchs pro Teil



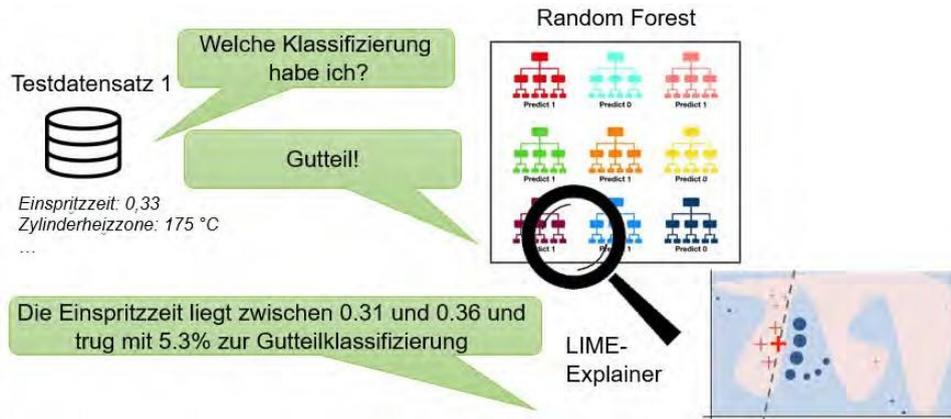
KI-Algorithmus Random Forest

- Gründe für die Auswahl von Random Forest:
 - Einzelne Entscheidungsbäume geben Gründe für Auswahl
 - Robust gegenüber fehlerhaften Daten
 - Interpretierbar
 - Schnelle Modellgenerierung und performanter Klassifizierungsprozess
 - Diskrete und kontinuierliche Klassifizierung
 - Geringe Überanpassung



KI-Algorithmus LIME

- Interpretierung des generierten Modells
- Erklärung der Klassifizierung



Gründe für Abweichungen in der Praxis

- Theoretischer Ansatz – Test von Machine Learning-Algorithmen für den vorgegebenen Einsatzzweck bezüglich der Relevanz für produzierende KMU
- Hochrechnung basiert auf Zeitraum von 4 Monaten, außerhalb des Zeitraums andere Energie durch andere Programme / Änderung der Außenbedingungen möglich
- Angabe der Energieeinsparung zwischen best case und worse case
- Korrelation zwischen Features
- Fehlermultiplikator: Energie nur auf 10Wh genau, addiert sich bei 12 Mio Teilen
- Reale Handlungsempfehlungen verbunden mit der Entwicklung einer einsatzfähigen Software im realen Produktionsbetrieb erfordern weit mehr personelle und finanzielle Kapazitäten.

Künstliches neuronales Netz

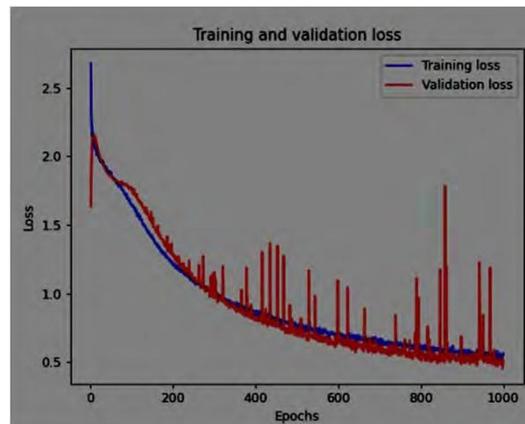
- Erweiterung des Prototyps um die Entwicklung eines künstlichen neuronalen Netzes
 - KNN werden für jede Maschine und für Qualität und Energie erstellt und optimiert.
 - Bei Maschinen mit großer Datengrundlage können gute Ergebnisse erzielt werden.

Künstliches neuronales Netz

- Erstellen der Modelle für Energie
- Finden lokaler Maxima der Modellfunktionen mittels Bayescher Optimierung → mögliche optimale Parameter
- Mathematisches Ermitteln der plausibleren Maxima
- Erklärung der Parameter mittels LIME

Künstliches neuronales Netz – Modelle (I)

M68 (8621 Daten- punkte)	Gute Qualität	Schlechte Qualität
Richtig erkannt	8670	73
Falsch erkannt	171 (2.0 %)	7 (8.7 %)



Künstliches neuronales Netz – Modelle (II)

- M68 mit Kamerdaten (137 Datenpunkte)

	Gute Qualität	Schlechte Qualität
Richtig erkannt	3	0
Falsch erkannt	0 (0 %)	134 (100 %)

Künstliches neuronales Netz – Modelle (II)

M69 (27529 Daten- punkte)	Gute Qualität	Schlechte Qualität
Richtig erkannt	23236	395
Falsch erkannt	3834 (14.2 %)	64 (13.9 %)

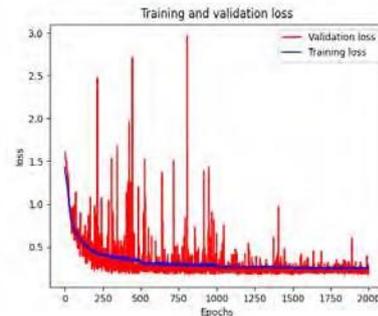
M67 (24706 Daten- punkte)	Gute Qualität	Schlechte Qualität
Richtig erkannt	22846	91
Falsch erkannt	1761 (7.2 %)	8 (8.0 %)

Erklärung der Fehlermatrix

- Fehler bedeutet nicht Fehlerrate bei der Herstellung, sondern Fehler des Modells bei der Kategorisierung von Schlechtteilen.
- Da möglichst viele Schlechtteile richtig erkannt werden sollen, muss man die Rate in rot (schlechte Qualität, falsch erkannt) minimieren.
- Modelle erzielen gute Ergebnisse, allerdings existiert noch keine Einbeziehung der Kameradaten.

Verfahren am Beispiel M64 (I)

- Nach Modelloptimierung:
 - Falsch erkannte Gute Qualität: 2,2%
 - Falsch erkannte Schlechte Qualität: 3,3%
- ➔ Gute Kategorisierung

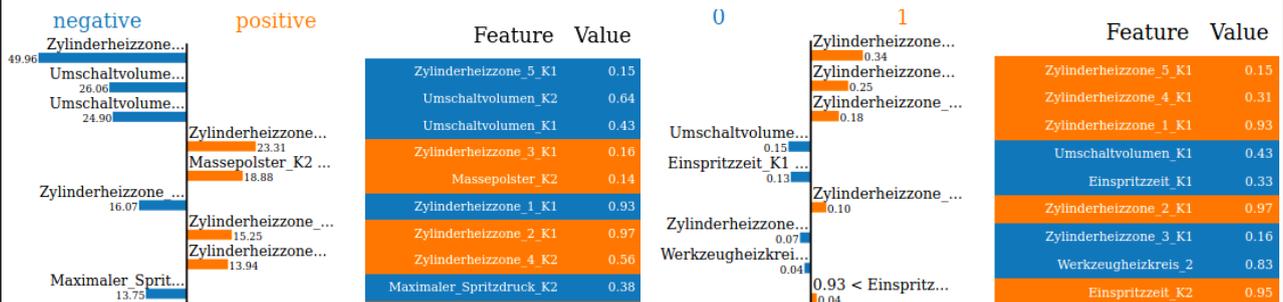


Verfahren am Beispiel M64 (II)

- Neuronales Netz hat eine Predict-Funktion, die aufgrund der gesehenen Werte eine Vorhersage für neue Werte gibt.
- Um Qualität und Energie einzubeziehen, werden diese in einem Maß vereint:
$$\text{predict}(p \text{ energy}, p \text{ quality}) = \text{predict}(p \text{ energy}) * (1 - \text{predict}(p \text{ quality}))$$
- Diese Predict-Funktion wird mittels der Hyperopt (fmin)-Funktion (Hyperparameteroptimierung) minimiert, um verschiedene lokale Minima zu erhalten.
- Dadurch soll die richtige Kombination von Parameterwerten gefunden werden, um in angemessener Zeit eine maximale Leistung für die Daten zu erzielen.

Verfahren am Beispiel M64 (III)

- Die verschiedenen Konfigurationen werden mit LIME erklärt:
- Energiemodell:
Qualitätsmodell:



Verfahren am Beispiel M64 (IIII)

- Die Parameter mit dem meisten Einfluss werden auf ihre einzelne Auswirkung in einer Probe von 1000 Werten getestet.
- Manche Parameter performen, wenn sie einzeln geändert werden, nicht mehr so gut.

Table 2: Quality and Energy Improvement on Sample

Parameter	Normalized Value	Quality Improvement in %	Energy Improvement in Wh
Zylinderheizzone_1_K1	0.93	6.65	-4.27
Zylinderheizzone_2_K1	0.97	4.87	6.29
Zylinderheizzone_4_K1	0.31	-11.44	13.26
Zylinderheizzone_5_K1	0.15	-1.36	48.89
Umschaltvolumen_K1	0.43	-36.90	-55.65
Umschaltvolumen_K2	0.64	-5.69	32.28
Einspritzzeit_K2	0.95	-4.69	-0.17
Maximaler_Spritzdruck_K2	0.38	-19.78	41.91

Verfahren am Beispiel M64 (V)

- Die 4 Parameter Zylinderheizzone 1 K1, Zylinderheizzone 2 K1, Zylinderheizzone 5 K1 und Umschaltvolumen K2 haben gut abgeschnitten.
- Im Anschluss wird eine abgeänderte Version des Standardfehlers des arithmetischen Mittels als Maß für die Vertrauenswürdigkeit des Parameters genommen.
- Alle 4 Parameter erhalten dafür gute Werte
- Somit stellen diese im Rahmen des Modells gute Änderungen dar.

Probleme Datenlage (I)

- Datenbank befinden sich auf Server von Novapax.
- 985.835 Datenpunkte
- Enthält eine Schuss_ID, Zeitstempel, Parameterwerte, Energiewerte, Qualitätswerte
- Allerdings: Qualitätswerte nicht auf Basis von Kameradaten
- Modelle s.o. auch ohne Kameradaten

Probleme Datenlage (II)

- Erhaltene Kameradaten in vielen .csv Dateien
- Keine Schuss_ID, nur Zeitstempel
- Daher: Programm zum Einlesen der Daten und Zuordnung zu Schuss_ID's in Datenbank auf Basis der Zeitstempel
- Allerdings stimmen nur 16103 der Zeitstempel überein.
- Viel zu wenig Daten für ein NN (verteilt auf 4 Maschinen)
- Kameradaten aus .csv Dateien fragwürdig, da es über 6 Millionen Datenpunkte aufweist, obwohl die Datenbank selber nur einen Bruchteil davon hat.

Lösungsansatz und nächste Schritte

- Programm zur Erhebung der Kameradaten und automatische Erfassung in der Datenbank statt .csv
- Sammeln von ca. 20.000 Datenpunkten je Maschine mit Kameradaten für die Qualität
- Nutzen des beschriebenen Verfahrens um Parameterempfehlungen zu bekommen

Nächste Schritte der Forschung/Anschlussarbeiten

- Überführung der Ergebnisse in die Praxis sowie Optimierung auf Basis der praktischen Umsetzung
 - Dazu kann für einige Wochen die normale Fertigung nicht stattfinden, daher war dieser Task bisher nicht möglich.
- Aufbau eines nationalen Kompetenznetzwerkes mit Best-Practices der bisherigen Ergebnisse
- Adaption der Ergebnisse auf andere produzierende Unternehmen und Überführung dieser Adaptionen in das Kompetenznetzwerk

Vielen Dank für die Aufmerksamkeit

