

Datengesteuertes maschinelles Lernmodell schätzt Effizienzgewinne durch passive Filter unter variablen Lasten

Uchenna Johnpaul Aniekwensi ^{a,*}, Dipyaman Basu ^b, Jörg Bausch ^a

^a Institut für Nachhaltige Energiesysteme (INES), Hochschule Offenburg, Badstraße 24, Offenburg 77652 Baden-Württemberg, Deutschland

^b Livarsa GmbH, Im Fruchtfeld 17, Berghaupten 77791 Baden-Württemberg, Deutschland

HIGHLIGHTS

- Ein zweistufiges ML-Modell sagt passive Filter Effizienz anhand von Feldmessungen.
- Nutzt reale Daten und überwindet Simulations Grenzen für standortspezifische Schätzungen.
- Ausgewählte Top-Merkmale durch Ridge-Regression, die von XGBoost zur Verbesserung der Vorhersage verwendet werden.
- Neutralstrom, Unsymmetrie, Oberschwingungen, Leistungsfaktor: wichtige Faktoren für Effizienzsteigerungen.
- Ermöglicht datengesteuerte Investitionen in „investment für Zuverlässiger Einsatz passiver Filter.

GRAFISCHE ZUSAMMENFASSUNG



ARTIKELINFORMATIONEN

Schlüsselwörter: Passive Filter
Energieeffizienz
Maschinelles Lernen
Stromqualität
Prädiktive Analytik
Datengesteuertes Modell

ZUSAMMENFASSUNG

Eine genaue Abschätzung der Stromverlustreduzierung durch passive Filter vor der Installation ist aufgrund variabler Lasten und Stromqualitätsbedingungen an den verschiedenen Netzpunkten schwierig. Bestehende Studien stützen sich auf Simulations- oder Analysemodelle. Diese Ansätze können die Variabilität in der Praxis mit datengesteuerten Methoden oft nicht erfassen. Diese Lücke schränkt effektive, standortspezifische Entscheidungen zum Einsatz von Filtern ein. Wir stellen einen zweistufigen Ansatz des maschinellen Lernens zur Abschätzung vor.

Energieeffizienzgewinne durch passive Filter unter variablen Bedingungen unter Verwendung hochauflösender Leistungsanalysedaten. Die Ridge-Regression identifiziert wichtige prädiktive Variablen und erreicht einen Basiswert von $R^2 = 0.591$. XGBoost erfasst dann nichtlineare Wechselwirkungen zwischen Lastschwankungen, Störungen der Stromqualität und Filterleistung und verbessert die Genauigkeit auf $R^2 = 0.755$. Die Methodik wurde in Zusammenarbeit mit der Livarsa GmbH durch den Einsatz in drei Industrieanlagen validiert. Die Ergebnisse zeigen einen durchschnittlichen relativen Fehler von 9,9 % über die gemessene Effizienz.

Gewinne, die die Zuverlässigkeit unter realen Bedingungen bestätigen. Umfassende Validierung durch k-fache Kreuzvalidierung, Ensemble-Methoden und externe Tests quantifizierten die Vorhersageunsicherheit, die kleinen industriellen Datensätzen (25 Trainingsbeispiele) innewohnt. Der Ansatz bietet ein skalierbares, datengestütztes Entscheidungshilfetooll, das simulationsbasierte Einschränkungen überwindet. Die Recheneffizienz ermöglicht eine Echtzeitbewertung während der Kundenberatung ohne spezielle Software. Der wirtschaftliche Wert ergibt sich aus reduzierten Leistungsgarantiespannen, beschleunigten Bewertungszeiträumen und minimierten Gewährleistungsrisiken. Zu den Einschränkungen zählen statistische Beschränkungen aufgrund begrenzter Trainingsdaten, die sich in einer Überanpassung der Kreuzvalidierung und breiten Konfidenzintervallen widerspiegeln. Die externe Validität erfordert eine standortspezifische Validierung für Anlagen mit erheblich unterschiedlichen elektrischen Eigenschaften. Trotz dieser Einschränkungen bieten die Ergebnisse einen praktischen Wert für Energieexperten, die nach effizienten Lösungen für die Stromqualität suchen, und ermöglichen zuverlässige Entscheidungen zum Einsatz passiver Filter auf der Grundlage quantifizierter Leistungsprognosen.

* Korrespondierender Autor.

E-Mail-Adresse: uchenna.aniekwensi@hs-offenburg.de (U.J. Aniekwensi).

1. Einleitung

Die genaue Abschätzung der Energieeffizienzvorteile passiver Filter vor der Installation bleibt eine Herausforderung. Diese Schwierigkeit ergibt sich aus den stark variierenden Lasten und Stromqualitätsbedingungen an verschiedenen Netzknoten. Probleme mit der Stromqualität, harmonische Verzerrungen, Spannungsungleichgewichte, Spannungsabfälle und -anstiege sowie ein hoher dynamischer Blindleistungsbedarf beeinflussen die Energieeffizienz und die Betriebssicherheit erheblich [1]. Der Übergang zu nichtlinearen, halbleiterbasierten Lasten hat diese Herausforderungen noch verschärft und zu erhöhten Energieverlusten und einer beeinträchtigten Zuverlässigkeit der Anlagen geführt [2].

Passive Filter bieten effiziente und kostengünstige Lösungen zur Verbesserung der Stromqualität in industriellen Anlagen. Sie minimieren Oberschwingungen und Verzerrungen, erhöhen den Leistungsfaktor und maximieren den wirtschaftlichen Nutzen durch direkte Energieeinsparungen, während sie gleichzeitig Schäden an den Anlagen verhindern [3]. Bei hohen Stromlasten sind sie oft die einzige praktikable Option zur Verbesserung der Stromqualität. Das Potenzial zur Effizienzsteigerung variiert jedoch stark zwischen den einzelnen Anlagen aufgrund der einzigartigen Lastcharakteristika, der Basisstromqualität und der Betriebsmuster in jeder Anlage.

Bestehende Ansätze liefern keine zuverlässigen standortspezifischen Vorhersagen. Die meisten Studien stützen sich auf Simulationsmodelle, um Effizienzgewinne durch THD-Reduzierung (Total Harmonic Distortion) zu schätzen [4] oder das Effizienzpotenzial durch Minimierung des Oberschwingungsverlustfaktors und Maximierung der Belastbarkeit zu quantifizieren [5]. Analytische Modelle wurden in Optimierungsalgorithmen integriert, um das Potenzial durch minimierte Wirkleistungsverluste zu schätzen [6]. Diese simulationsbasierten und analytischen Methoden weisen entscheidende Einschränkungen auf: Sie spiegeln die Variabilität in der Praxis nicht genau wider und messen nicht die standortspezifischen Energieeffizienzgewinne unter tatsächlichen Betriebsbedingungen.

Diese Lücke führt zu erheblichen kommerziellen und technischen Herausforderungen. Anbieter passiver Filter sind mit Risiken hinsichtlich der Leistungsgarantie und potenziellen Reputationsschäden aufgrund ungenauer Vorhersagen konfrontiert. Systementwickler, Betreiber und Endnutzer benötigen zuverlässige Filterlösungen, die für ihre individuellen Umgebungen optimiert sind, doch herkömmliche Ansätze können diesen Anforderungen nicht gerecht werden. Der Mangel an auf die Betriebsdynamik zugeschnittenen Prognosetools schränkt effektive Entscheidungen zum Filtereinsatz und das Vertrauen in die erwarteten Kapitalrenditen ein.

Die entscheidende Forschungslücke besteht im Fehlen datengestützter Prognosetools, die die realen Betriebsvariabilitäten für die standortspezifische Leistungsabschätzung passiver Filter erfassen. Simulationsmodelle liefern zwar theoretische Erkenntnisse, können jedoch die komplexen Wechselwirkungen zwischen variablen Lasten, Störungen der Stromqualität und Filterleistung unter tatsächlichen Betriebsbedingungen nicht berücksichtigen. Datengestützte Ansätze sind für das Management der dynamischen Komplexität moderner elektrischer Systeme unerlässlich. Studien zur Optimierung intelligenter Stromnetze bestätigen, dass prädiktive Modellierung zu quantifizierbaren Verbesserungen der Effizienz und einer Verringerung der Spitzenlast führt [7,8]. Die Verbesserung der Energieeffizienz durch genaue Vorhersagen ist für moderne Stromversorgungssysteme von entscheidender Bedeutung, um eine nachhaltige Entwicklung zu gewährleisten und gleichzeitig die Widerstandsfähigkeit und Zuverlässigkeit des Netzes zu erhalten. Wir stellen einen datengesteuerten Ansatz des maschinellen Lernens vor, der diese Lücke schließt, indem er die Effizienzgewinne passiver Filter anhand hochauflösender Daten von Leistungsanalysatoren, die aus Betriebsanlagen gesammelt wurden, genau abschätzt.

Unsere zweistufige Methodik verwendet Ridge-Regression für die Rangfolge der Merkmalsbedeutung und die Dimensionsreduktion, gefolgt von XGBoost für die nichtlineare Vorhersage. Dieser Ansatz schätzt die Effizienzgewinne unter variablen Lasten und Stromqualitätsbedingungen präzise ab. Bedingungen, die durch Simulation nicht erfasst werden können.

Die Methodik umfasst ein gründliches Feature Engineering anhand aktueller Wellenformen, Effektivwerte und aggregierter Stromqualitätsindizes aus realen Betriebsdaten. Die Ridge-Regression behandelt die Multikollinearität zwischen den Stromqualitätsvariablen und bewertet die Bedeutung der Merkmale anhand regulierter Koeffizienten, wodurch interpretierbare Basisprognosen erstellt werden. XGBoost erfasst dann komplexe nichtlineare Beziehungen zwischen Lastschwankungen, Störungen der Stromqualität und Filterleistung, die lineare Modelle nicht darstellen können.

Unsere Arbeit baut auf der Erweiterung der Forschung zur Anwendung von maschinellem Lernen auf die Vorhersage von Energieeinsparungen und die Analyse der Stromqualität auf [9,10]. Während frühere Studien Ridge-Regression für Energieanwendungen [11, 12] und XGBoost für Energieprognosen [13,14] untersuchten, zielt unser spezifischer Beitrag auf Effizienzsteigerungen durch passive Filter in Umgebungen mit variabler Stromqualität unter Verwendung von Feldmessdaten ab. Damit wird eine kritische Lücke zwischen theoretischer Modellierung und praktischer Energiesystemtechnik geschlossen, wo operative Variabilität das Systemverhalten dominiert.

Die praktischen Auswirkungen sind erheblich. Der Ansatz bietet ein skalierbares Tool, das datengestützte Investitionsentscheidungen zur Verbesserung der Stromqualität unterstützt. Durch die Ermöglichung genauer, standortspezifischer Vorhersagen unter realen Schwankungen verbessert es die Planung passiver Filter und die Zuverlässigkeit ihres Einsatzes. Die Recheneffizienz ermöglicht eine Echtzeitbewertung während der Kundenberatung ohne spezielle Simulationssoftware. Der wirtschaftliche Wert ergibt sich aus reduzierten Leistungsgarantiespannen, einer Beschleunigung der Bewertungszeiträume von Wochen auf Minuten und einer Minimierung des Garantieaufwands.

Die transparente Anerkennung von Einschränkungen des Umfangs stärkt das Vertrauen in die gemeldeten Fähigkeiten. Der Trainingsdatensatz mit 25 Stichproben unterliegt statistischen Einschränkungen, die sich in einer Überanpassung bei der Kreuzvalidierung und großen Konfidenzintervallen widerspiegeln. Die externe Validität erfordert eine standortspezifische Validierung für Anlagen mit erheblich unterschiedlichen elektrischen Eigenschaften. Diese Einschränkungen spiegeln grundlegende Herausforderungen bei der industriellen Datenerfassung wider, bei der praktische Hindernisse wie wirtschaftliche Belastungen, begrenzte Verfügbarkeit von Partnern und Sicherheitsbeschränkungen eine umfangreiche Erweiterung des Datensatzes verhindern. Einschränkungen hinsichtlich der Datenverfügbarkeit [15] und der Modellsensitivität [16] schränken die Verallgemeinerungsfähigkeit ein und deuten auf eine mögliche Variabilität der Ergebnisse hin.

Trotz dieser Einschränkungen stellt diese Arbeit einen bedeutenden Fortschritt bei der Anwendung von maschinellem Lernen für pragmatische Lösungen zur Stromqualitätsverbesserung dar. Die Forschung wurde gemeinsam mit dem Industriepartner Livarsa GmbH durchgeführt, der Zugang zu Betriebsdaten aus 28 Anlagen gewährte und die Validierung vor Ort durch den Einsatz an drei Teststandorten ermöglichte. Dieser Artikel beschreibt detailliert die Methodik (Datenerfassung, Feature Engineering, Modellentwicklung und Validierung) und die Validierungsergebnisse und demonstriert die Wirksamkeit von Ridge Regression und XGBoost bei der Vorhersage von Energieeinsparungen, wobei Einschränkungen und Grenzen des Anwendungsbereichs transparent dargelegt werden.

2. Verwandte Arbeiten

Die steigende Nachfrage nach Energieeffizienz und zuverlässiger Stromqualität hat zu umfangreichen Forschungen im Bereich passiver Filteranwendungen und zur Entwicklung von Vorhersagemodellen geführt [17]. In diesem Abschnitt werden relevante Literatur zu den Themen Verbesserung der Stromqualität, passive Filteranwendungen und maschinelle Lernverfahren für die Energievorhersage untersucht, wobei der Schwerpunkt auf der Bewältigung der Herausforderungen liegt, die sich aus der Variabilität kundenspezifischer Stromqualitätsszenarien ergeben.

2.1. Stromqualität und passive Filteranwendungen

Die Bedeutung der Aufrechterhaltung der Stromqualität innerhalb akzeptabler Grenzen, wie sie in Normen wie IEEE 519, IEC 61000-3-6 und IEC 50160 definiert sind, wurde umfassend untersucht. Oberschwingungsverzerrungen, Spannungseinbrüche/-anstiege und Probleme mit dem Leistungsfaktor können industrielle Anlagen und die Energieeffizienz erheblich beeinträchtigen [18]. Passive Filter bieten bewährte Lösungen zur Reduzierung von harmonischen Verzerrungen und zur Verbesserung des Leistungsfaktors, einschließlich Anwendungen in dreiphasigen netzgekoppelten erneuerbaren Energiesystemen [19]. Es wurden verschiedene Entwurfs- und Optimierungsstrategien für Filter vorgeschlagen, die sich auf Abstimmungsmethoden und Resonanzunterdrückung konzentrieren, um eine maximale harmonische Dämpfung zu erreichen [20,21]. Die Forschung zur Theorie der Momentanleistung beeinflusst auch die Konstruktion und Steuerung von Leistungsfiltern [22]. Während traditionelle Studien oft die theoretischen Vorteile passiver Filter hervorheben, werden praktische Leistungsgarantien auf der Grundlage empirischer, standortspezifischer Daten in der Literatur selten behandelt, was die Notwendigkeit der in diesem Beitrag vorgestellten Vorhersagewerkzeuge unterstreicht.

Arbeit.

2.2. Maschinelles Lernen für Energieprognosen und Regressionsanalysen

Maschinelle Lernverfahren, insbesondere Regressionsmodelle, werden zunehmend für die Energieprognose eingesetzt. Übersichtsarbeiten wie [23] heben das Potenzial von Regressionsmodellen hervor, komplexe Zusammenhänge für die Prognose des Energieverbrauchs von Gebäuden zu erfassen. Insbesondere die Ridge-Regression ist wertvoll für den Umgang mit Multikollinearität und die Vermeidung von Überanpassung, die häufige Probleme bei der Analyse von Stromqualitätsdaten sind. Ihre theoretischen Grundlagen und praktischen Anwendungen werden in [24] diskutiert. Die Leistungsfähigkeit dieser Modelle hängt jedoch häufig von der Verfügbarkeit ausreichender Daten und der Stabilität der Trainingsparameter ab.

Über den Umgang mit Multikollinearität und die Vermeidung von Überanpassung mithilfe von Methoden wie der Ridge-Regression hinaus bieten Algorithmen des maschinellen Lernens leistungsstarke und vielversprechende Lösungen für die Energieeffizienz in komplexen intelligenten Netzen, indem sie Nutzungsmuster vorhersagen, Ineffizienzen identifizieren und die vorausschauende Wartung unterstützen. Der nachgewiesene Erfolg der prädiktiven Modellierung und Optimierung unter Verwendung von Algorithmen wie Regression, Clustering und neuronalen Netzen bei der Erzielung erheblicher Vorteile wie 15

Eine Verbesserung der Energieeffizienz um 25 % und eine Reduzierung des Spitzenbedarfs um 25 % unterstreichen den nachgewiesenen Nutzen von ML in industriellen Energiesystemen [7].

Während etablierte Regressionsmodelle wie Ridge Regression für den Umgang mit Multikollinearität und die Vermeidung von Überanpassung in Stromqualitätsdaten wertvoll sind, erfordert die Anwendung von maschinellem Lernen in dezentralen oder ressourcenbeschränkten Umgebungen zunehmend leichte, skalierbare und Echtzeit-Lösungen. So entstehen beispielsweise neuartige Frameworks, die Property Testing in Kombination mit Edge AI und Federated Learning nutzen, um Anomalien in komplexen verteilten Netzwerken (z. B. IoT) effizient zu erkennen und zu klassifizieren, wodurch sich die Verarbeitungszeit und der Energieverbrauch erheblich reduzieren lassen, während gleichzeitig eine hohe Genauigkeit gewährleistet bleibt [25].

2.3. Baumbasierte Ensemble-Methoden: XGBoost und darüber hinaus

Über traditionelle Regressionsmodelle wie Ridge Regression hinaus haben baumbasierte Ensemble-Methoden wie Gradient Boosting und dessen optimierte Implementierung, XGBoost [26], bemerkenswerte Erfolge bei verschiedenen Vorhersageaufgaben erzielt, darunter auch bei der Energieprognose [13,14]. Diese Methoden erstellen ein starkes Vorhersagemodell, indem sie die Ergebnisse mehrerer Entscheidungsbäume kombinieren, wobei jeder Baum sequenziell aus den Fehlern seiner Vorgänger lernt. Insbesondere XGBoost integriert Regularisierungstechniken und effiziente Algorithmen zur Baumbildung, was oft zu einer überlegenen Leistung im Vergleich zu traditionellen linearen Modellen führt, insbesondere bei komplexen, nichtlinearen Beziehungen in den Daten. Studien haben gezeigt, dass XGBoost komplexe Muster in Energieverbrauchsdaten effektiv erfassen und eine hohe Vorhersagegenauigkeit erzielen kann [27]. Seine Fähigkeit, Nichtlinearitäten und Merkmalsinteraktionen zu verarbeiten, macht es zu einem vielversprechenden Kandidaten für die Modellierung der komplexen Beziehung zwischen Stromqualitätsparametern und Effizienzgewinnen. Darüber hinaus bietet XGBoost Mechanismen zur Bewertung der Merkmalsbedeutung und liefert Einblicke in die Treiber der vorhergesagten Ergebnisse [26]. Während wir uns in der Anfangsphase unserer Arbeit auf das lineare Ridge-Regressionsmodell konzentrierten, um eine Basislinie zu erstellen und Multikollinearität zu berücksichtigen, motivierte uns das Potenzial, komplexere Beziehungen zu erfassen und eine höhere Vorhersagegenauigkeit zu erzielen, dazu, XGBoost als alternativen Modellierungsansatz zu untersuchen. Diese Verlagerung steht im Einklang mit dem allgemeinen Trend im maschinellen Lernen, leistungsstarke Ensemble-Methoden für anspruchsvolle Vorhersageprobleme einzusetzen.

Die Suche nach optimalen Merkmalsätzen für komplexe Vorhersageaufgaben hat zur Einführung fortschrittlicher metaheuristischer Optimierungsalgorithmen geführt. Diese Optimierungsmethoden werden häufig für die binäre Merkmalsauswahl eingesetzt, um die Vorhersagegenauigkeit zu verbessern und gleichzeitig die Komplexität der Merkmale zu minimieren. So haben beispielsweise spezielle Algorithmen wie die Adaptive Dynamic Dipper Throated Optimization (bGW-DTO) eine überlegene Fähigkeit bei der Identifizierung minimaler, hochwirksamer

Merkmalsuntergruppen für Klassifizierungsprobleme zu identifizieren, wobei sie oft eine sehr hohe Vorhersagegenauigkeit und Verarbeitungsgeschwindigkeit erreichen. Solche ausgefeilten Optimierungsansätze stehen im Gegensatz zu Regularisierungstechniken wie der Ridge-Regression und zeigen den wachsenden Nutzen der Merkmalsauswahl in verschiedenen Bereichen wie der Energievorhersage und der medizinischen Diagnose [28].

Die Wirksamkeit des maschinellen Lernens bei komplexen, nichtlinearen Vorhersageproblemen hängt oft von einer robusten Modellstruktur und einer Hyperparameter-Optimierung ab. Über allgemeine Ensemble-Ansätze hinausgehend nutzen spezialisierte Methoden fortschrittliche metaheuristische Optimierungsalgorithmen wie den hybriden PSOBER-Algorithmus (Particle Swarm Optimization and Al-Biruni Earth Radius Optimization), um komplexe Netzwerkparameter (wie diejenigen in CNNs und DBNs) für Aufgaben zu optimieren, die eine hohe Vorhersagegenauigkeit erfordern. Diese Optimierungstechniken sind entscheidend für die Maximierung der Genauigkeit und die Minimierung von Klassifizierungsfehlern in Szenarien, die durch eine hohe Datenvariabilität gekennzeichnet sind [29].

Die Komplexität und Dimensionalität von Daten in industriellen und Netzwerküberwachungsbereichen erfordern eine strenge Merkmalsauswahl, insbesondere wenn sich herkömmliche lineare Methoden als unzureichend erweisen. Folglich werden hybride metaheuristische Algorithmen wie die GWDTO-Methode (Grey Wolf and Dipper Throated Optimization) zunehmend eingesetzt, um das Problem der Merkmalsauswahl (FS) in hochdimensionalen Bereichen wie IoT-Netzwerk-Intrusion-Detection-Systemen (IDS) zu lösen. Diese Methoden wurden entwickelt, um auf effiziente Weise minimale, optimale Teilmengen von Merkmalen zu identifizieren, die die Klassifizierungsgenauigkeit maximieren, indem sie das Gleichgewicht zwischen den Explorations- und Auswertungsphasen des Optimierungsprozesses verbessern. Dieser Ansatz bestätigt die allgemeine Notwendigkeit einer robusten Merkmalsauswahl vor einer hochgenauen Vorhersage in verschiedenen technischen Bereichen [30].

Darüber hinaus stützt sich die Optimierung komplexer Ensemble-Modelle häufig auf fortschrittliche Techniken, die über herkömmliche Hyperparameter-Gitter hinausgehen. Hybride metaheuristische Optimierungsalgorithmen, wie beispielsweise der Adaptive Dynamic Sine Cosine Fitness Grey Wolf Optimization (ADSCFGWO)-Algorithmus, wurden erfolgreich eingesetzt, um die Gewichte und Parameter von Multi-Modell-Voting-Klassifikatoren (NN, SVM, KNN) zu optimieren und so die Genauigkeit bei komplexen nichtlinearen Vorhersageaufgaben wie der Unkrautdetektion in Drohnenbildern deutlich zu verbessern. Der Einsatz solcher Optimierungsmethoden bestätigt den breiteren Nutzen von Metaheuristiken zur Verbesserung der Ensemble-Lernleistung in kritischen datengesteuerten Anwendungen [31].

2.4. Umgang mit Datenbeschränkungen und Modellstabilität

Eine entscheidende Einschränkung unserer Studie ist die begrenzte Anzahl von Datenproben, die die Verallgemeinerungsfähigkeit des Modells einschränken kann. Die Forschung zur Regression bei kleinen Stichprobengrößen [15] betont die Herausforderungen und potenziellen Verzerrungen, die mit begrenzten Datenmengen verbunden sind. Über die Beschränkungen der Stichprobengröße hinaus ist die Sicherstellung der externen Validität, also die Bestätigung, dass die vorgeschlagene Methodik auch außerhalb des ursprünglichen Validierungssatzes an verschiedenen Industriestandorten wirksam bleibt, eine wiederkehrende Einschränkung, die in systematischen Literaturübersichten zu Frameworks für maschinelles Lernen häufig angesprochen wird. Techniken zur Minderung dieser Probleme, wie Regularisierung und Kreuzvalidierung, sind für die Gewährleistung der Modellrobustheit von entscheidender Bedeutung [32].

Darüber hinaus unterstreicht die beobachtete Empfindlichkeit unseres mit Ridge Regression trainierten Modells gegenüber dem Zufallsstartwert das Problem der Modellstabilität. Studien zur Empfindlichkeit gegenüber Zufallsstartwerten in maschinellen Lernmodellen [15, 16] untersuchen den Einfluss von Initialisierungsparametern auf die Modellleistung. Techniken wie Ensemble Learning und robuste Optimierung können diese Einschränkung möglicherweise beheben.

2.5. Neuartigkeit und Beitrag

Unsere Arbeit leistet einen bedeutenden Beitrag zu diesem Forschungsgebiet, indem sie einen datengesteuerten, zweistufigen maschinellen Lernansatz einführt, der speziell auf die genaue Abschätzung der Leistungsreduzierung durch passive Filter in Umgebungen mit variabler Stromqualität zugeschnitten ist. Damit wird direkt auf die entscheidende kommerzielle und technische Herausforderung eingegangen, die Filterleistung unter realen, dynamischen Bedingungen zu gewährleisten, was mit herkömmlichen Simulations- oder

Analytische Modelle haben dies weitgehend außer Acht gelassen. Während frühere Studien die Verwendung von Regressionsmodellen für die Energieprognose untersucht haben, konzentriert sich unser Ansatz in einzigartiger Weise auf die Quantifizierung von Effizienzsteigerungen passiver Filter unter Verwendung hochauflösender Leistungsanalysedaten, einschließlich rigoroser Feature-Engineering, und bietet eine empirische Validierung mit Schwerpunkt auf linearer (Ridge-Regression für Basislinie und Unsicherheit) und nichtlinearer (XGBoost für überlegene Genauigkeit) Modellierung. Insbesondere die Anwendung von XGBoost stellt einen bemerkenswerten Fortschritt bei der Erfassung der komplexen, nichtlinearen Wechselwirkungen dar, die den Stromqualitätsdaten innewohnen, und bietet ein robustes und skalierbares Werkzeug für datengestützte Investitionsentscheidungen.

Diese Neuerung, die erfolgreich die Lücke zwischen simulationsbasierten Studien und realitätsnahen Vorhersagemodellen für Energiesysteme schließt, zeigt, wie groß der praktische Nutzen sein kann, wenn fortschrittliches ML (Machine Learning) auf spezifische Herausforderungen in bestimmten Bereichen angewendet wird, ähnlich wie bei komplexen diagnostischen Problemen wie der Vorhersage von Hepatitis C im Gesundheitswesen [8]. Diese Vorhersageergebnisse in Energiesystemen unterstreichen das Potenzial von ML für eine verbesserte Entscheidungsfindung in der Industrie und eine höhere Wirtschaftlichkeit.

Während unsere Methodik Ridge-Regression für die Merkmalsauswahl und die Festlegung linearer Basislinien verwendet, werden in der Datenwissenschaft derzeit zunehmend metaheuristische binäre Optimierungsmethoden (z. B. bADSCFGWO) eingesetzt, um optimale Merkmalsuntergruppen zu identifizieren und so eine fortschrittliche Merkmalsauswahl mit optimierten Ensemble-Vorhersagemodellen zu kombinieren, um eine überlegene Gesamtleistung und statistische Signifikanz zu erzielen. Diese Verlagerung hin zur dualen Optimierung bestätigt die Notwendigkeit einer strengen Merkmalsauswahl vor der nichtlinearen Vorhersage in verschiedenen wissenschaftlichen Bereichen [31].

Unsere zweistufige Methodik, bei der Ridge-Regression für eine zuverlässige Merkmalsauswahl und anschließend XGBoost für die Erfassung komplexer nichtlinearer Wechselwirkungen zum Einsatz kommen, geht direkt auf die Herausforderung ein, die Leistung passiver Filter unter dynamischen, realen Bedingungen zu gewährleisten. Diese Betonung einer effizienten Merkmalsextraktion und einer robusten nichtlinearen Vorhersage ist von entscheidender Bedeutung, um Ressourcen- und Datenbeschränkungen zu überwinden. Die Notwendigkeit der Ressourceneffizienz wird durch innovative Lösungen in verwandten Bereichen noch weiter unterstrichen, wie beispielsweise das Property Testing und Federated Edge AI (Artificial Intelligence) Framework, das eine sublineare Anomalieerkennung in eingeschränkten Netzwerken ermöglicht und den allgemeinen Trend zu rechnerisch sparsamen, aber dennoch genauen datengesteuerten Entscheidungshilfetoools verdeutlicht [25].

3. Methodik

In diesem Abschnitt wird die Methodik zur Entwicklung eines prädiktiven Machine-Learning-Modells zur Abschätzung der durch die Installation eines passiven Filters erzielbaren Stromverlustreduzierungen beschrieben. Der Prozess umfasst die Datenerfassung aus verschiedenen Anlagen, die Datenverarbeitung, ein rigoroses Feature Engineering sowie die Auswahl und das Training der Algorithmen Ridge Regression und XGBoost. Das übergeordnete Ziel des Modells ist es, die potenziellen Effizienzverluste für Anlagen vor der Installation des Filters vorherzusagen. Diese Vorhersage dient als wichtige Grundlage für die Bewertung der Wirtschaftlichkeit der Filterbeschaffung auf der Grundlage der bestehenden Stromqualitätsbedingungen der Anlage. Abb. 1 zeigt eine visuelle Darstellung des Datenflusses während der Datenerfassung für das Modelltraining und dessen anschließende Anwendung für eine neue Anlage. Um das Verlustreduktionspotenzial des Filters für eine neue Anlage genau einschätzen zu können, wurde das Modell mit Daten trainiert, die bei der Umgehung des passiven Filters in bestehenden Anlagen erfasst wurden. Die spezifische Methode zur Quantifizierung der Effizienzgewinne (ECV®) wird in Abschnitt 3.2.1 detailliert beschrieben.

3.1. Datenerfassung

Der Datenerfassungsprozess umfasste einen Messaufbau, der an verschiedenen Standorten der Anlage eingesetzt wurde und aus einem Leistungsanalysator [34] und einem passiven Filter [35] bestand. Kritische elektrische Daten, die Aufschluss über die Stromqualität geben, wurden mit unterschiedlichen Zeitaufösungen aufgezeichnet. Insbesondere

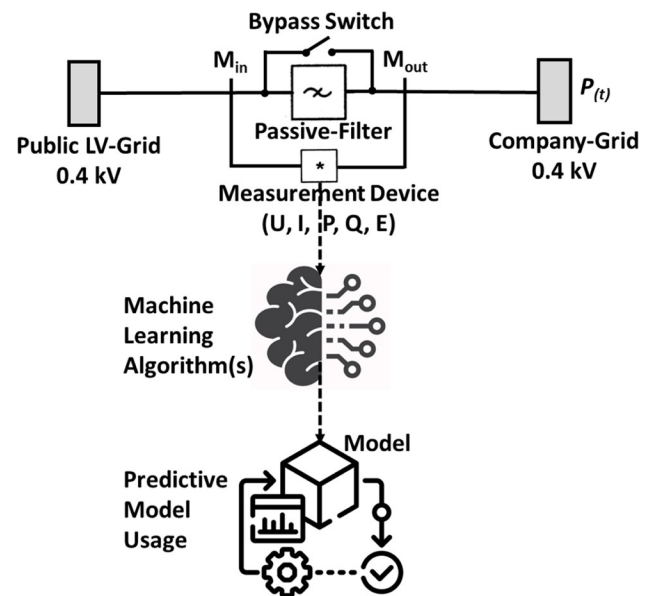


Abb. 1. Datenerfassung und Modellverwendungs-Workflow. Ein passiver Filter mit Bypass-Schalter wird zwischen dem öffentlichen Niederspannungstransformator (0,4 kV) und dem Unternehmensnetz (0,4 kV) installiert. Die mit einem Leistungsanalysator erfassten und gesammelten Daten werden vor der Schulung eines maschinellen Lernmodells vorverarbeitet und feature-engineered, das dann für die Entscheidungsfindung eingesetzt wird [33].

Die Abschaltung des passiven Filters während der Messkampagne löste ein vorübergehendes Ereignis aus, das automatisch aufgezeichnet und in der Webschnittstelle des Leistungsanalysators (zugänglich über die IP-Adresse des Geräts) gespeichert wurde. Die Metadaten dieses Ereignisses wurden in den Dateiformaten PQDIF (Power Quality Data Interchange Format) und Comtrade (Konfiguration und Daten) gespeichert.

Die PQDIF-Dateien (gemäß IEEE Std 1159.3) dienen als primäre Quelle für die Extraktion von zwei unterschiedlichen Daten mit unterschiedlichen Zeitaufösungen unter Verwendung der PQDiffactor-Software [36]:

- 10-Millisekunden-Effektivwerte: Stellen die Halbzyklus-Effektivwerte von Strom, Spannung usw. dar und geben Aufschluss über die unmittelbaren Auswirkungen des passiven Filter-Bypass-Ereignisses.
- 40-Mikrosekunden-Momentanwerte: Erfassen die detaillierten Wellenformcharakteristika, die für die Analyse des transienten Verhaltens und des Oberwellengehalts entscheidend sind.

Die Abbildungen 2 und 3 veranschaulichen diese beiden unterschiedlichen Zeitaufösungen. Zusätzlich wurden 2-Sekunden-Aggregationsdaten [37], die während des ECV®-Messzeitraums zentral erfasst und gespeichert wurden, verwendet, um einen umfassenderen Überblick über die Stromqualitätsbedingungen im Zeitverlauf zu erhalten.

3.1.1. Zusammensetzung des Datensatzes und industrielle Vielfalt

Der Datensatz umfasst 28 Industrieanlagen aus dem Kundenportfolio der Livarsa GmbH in Deutschland und der Schweiz, die sieben verschiedene Fertigungsbranchen repräsentieren: Kunststoffverpackungen, Aluminiumsandguss, Verschluss-technik, Wellpappenverpackungen, Bodenfliesenherstellung, technische Kunststoffsysteme (Teile, Baugruppen und technische Systeme) und Baustoffproduktion (Innenausbau und Dämmung). Diese sektorale Vielfalt gewährleistet die Darstellung unterschiedlicher Oberschwingungsprofile, Lastcharakteristiken und Stromqualitätsbedingungen, die für industrielle Anwendungen typisch sind.

Die Filterinstallationskapazitäten reichen von 800 A bis 3200 A, was einer vierfachen Variation der elektrischen Infrastruktur entspricht. Dieser Bereich umfasst sowohl mittelgroße Fertigungsbetriebe als auch große Industrieanlagen und bietet Anwendungsmöglichkeiten für typische Szenarien des Einsatzes passiver Filter in gewerblichen Anwendungen. Die Messungen jeder Anlage umfassen umfassende Daten zur Stromqualität, die während des normalen Betriebs erfasst wurden.

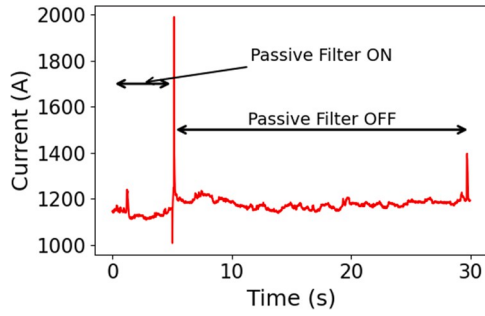


Abb. 2: Langfristige RMS-Stromreaktion auf das Umschalten des Filters. Der RMS-Strom wird über einen Zeitraum von 30 s im Verhältnis zur Zeit dargestellt. Eine große Stromtransiente tritt bei $t \approx 5$ s auf, wenn der passive Filter von EIN auf AUS umgeschaltet wird.

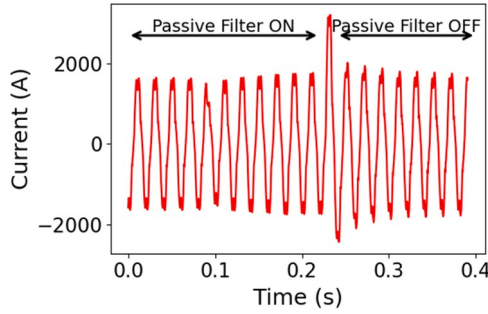


Abb. 3: Wechselstromwellenform während des Filterwechsels. Der Strom wird über einen Zeitraum von 400 ms dargestellt. Der passive Filter ist zunächst eingeschaltet

und wird bei $t \approx 0,23$ s ausgeschaltet, was zu einer großen transienten Spitze führt, gefolgt vom fortgesetzten, ungefilterten Wechselstrom.

Bedingungen, die realistische Lastschwankungen und Oberschwingungsanteile erfassen.

Während die absolute Stichprobengröße von 28 die praktischen Einschränkungen bei der industriellen Datenerhebung widerspiegelt, bietet die sektorale und skalare Vielfalt des Datensatzes eine breitere Repräsentativität als Labor- oder simulationsbasierte Studien mit homogenen Bedingungen. Das 25:3-Verhältnis von Training zu Test split behält diese Vielfalt über Validierungspartitionen hinweg bei.

3.2. Feature Engineering

Der Feature-Engineering-Prozess umfasste die Extraktion und Transformation relevanter Informationen aus den oben diskutierten Rohdatenquellen, um einen umfassenden Datensatz für das Modelltraining zu erstellen. Dies führte zu insgesamt 47 Merkmalsvariablen, und die Zielvariable (prozentuale Effizienzgewinne) wurde mit der im nächsten Abschnitt beschriebenen Methode berechnet.

3.2.1. Zielvariable: Daten zu Effizienzsteigerungen

Die Zielvariable, die die durch den passiven Filter erzielten Effizienzgewinne darstellt, wurde unter Verwendung der ECV®-Messmethode (Energy Comparison Value) quantifiziert [38]. Diese Methode (Abb. 4) wurde entwickelt, um eine überprüfbare Messung der Energieeffizienz in komplexen elektrischen Netzwerken zu ermöglichen, wobei die ungleichmäßigen Energieverbrauchsmuster aufgrund schwankender Lasten und Betriebsbedingungen berücksichtigt werden. Die ECV®-Methode ermöglicht einen aussagekräftigen Vergleich des Energieverbrauchs mit und ohne Filterbetrieb.

3.2.2. Merkmalsvariablen

Die Merkmalsvariablen wurden aus den drei primären Datenquellen (10-Millisekunden-Effektivwerte, 40-Mikrosekunden-Momentanwerte und 2-Sekunden-Aggregatdaten) abgeleitet.

3.2.2.1. Effektive Stromdaten (Auflösung 10 Millisekunden). Während der 24-

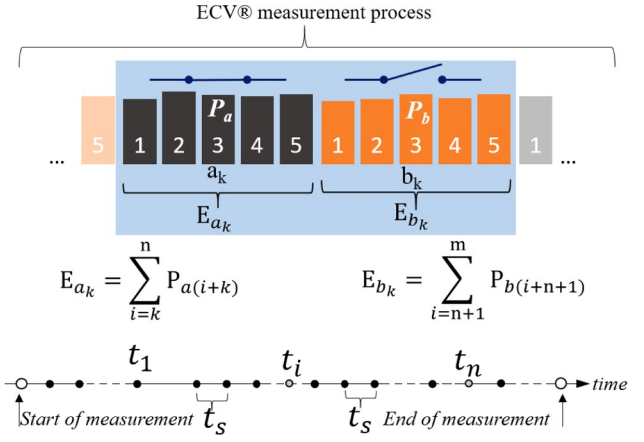


Abb. 4: Darstellung der Bewertung des Energievergleichswerts. Intervalle der Energiemessungen, wenn sich der passive Filter im Spar-/Bypass-Modus befindet. Die beigefügten Formeln definieren, wie der Energieverbrauch für jedes Intervall berechnet wird, indem die entsprechenden aufgezeichneten Leistungsmessdaten über die jeweiligen Messschritte summiert werden [38].

Bei der stündlichen ECV®-Messung werden einige PQDIF-Dateien (in der Regel eine Datei pro Stunde während der Hauptbetriebszeiten der Anlage) aufgezeichnet. Ein gemittelter Wert der abgeleiteten Variablen (z. B. $S_{(p)}$, Abb. 5) wurde berechnet, indem die Werte aller aufgezeichneten Beobachtungen in jeder der während des ECV®-Messzeitraums gesammelten PQDIF-Dateien unter Verwendung von (1) gemittelt wurden. Nach einer ähnlichen Methodik wie bei der Berechnung von $S_{(p)}$ wurden andere grundlegende Variablen aus den effektiven Stromdaten unter Verwendung der Gleichungen (2)-(6) berechnet.

$$S_p = S_{\text{Niles} \times \text{Nsamples}} \left[A(i, t) \right]_p \quad (1)$$

$$D_{p-q} = S_{\text{Niles} \times \text{Nstichproben}} \left[A, a(i, t) \right] \quad (2)$$

$$a(A)(i, t) = \frac{1}{3} \sum_{p=1}^3 A_p(i, t) \quad (3)$$

$$\mu = S_{\text{Niles} \times \text{Nstichproben}} \left[a(A)(i, t) \right] \quad (4)$$

$$DN_p = S_p / \mu * 1000 \quad (5)$$

$$DN_{p-q} = S_{p-q} / \mu * 1000 \quad (6)$$

$$DN_{\underline{N}_p} = a(DN_p) / \mu * 1000 \quad (7)$$

$$DN_{\underline{N}_{p-q}} = a(DN_{p-q}) / \mu * 1000 \quad (8)$$

wobei

Indizes:

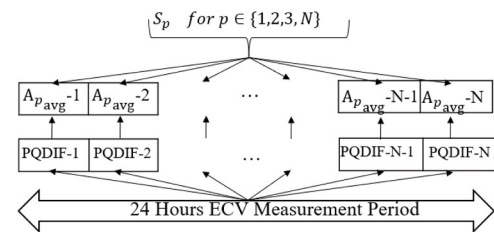


Abb. 5: Feature Engineering-Ablauf für die Modelleingabe. Der Prozess veranschaulicht, wie die während eines 24-stündigen ECV-Messzeitraums gesammelten Daten transformiert werden. Rohdaten aus PQDIF-Dateien (PQDIF-1 bis PQDIF-N einer bestimmten Anlage) werden verwendet, um die Berechnung des gemusterten Mittelwerts ($A_{p_{avg}-1}$ bis $A_{p_{avg}-N}$) verwendet, der dann aggregiert wird, um den endgültigen Satz von Merkmalsvariablen (z. B. S_p) für das Modell zu erstellen.

i : Index für Proben innerhalb des Messtages.

t : Index für Proben innerhalb einer Datei.

$p - q$: Phasenindizes (typischerweise $p - q \in \{1, 2, 3\}$ für ein dreiphasiges System).

Datensätze und Aggregationsnotation:

$S_{ij}^{N_{Dateien}, N_{Stichproben}}$ [.] – gemittelter Stichprobenwert, der auf eine Variable über $N_{Dateien}$ und $N_{Stichproben}$ -Zeitchstichproben angewendet wird.

Variablen:

S_p : gemittelter Wert der Variablen A_p für die Phase p über alle Dateien und Zeitabstastungen.

D_{p-q} : Abgetasteter Mittelwert der Kreuzphasenbeziehung zwischen den Phasen p und q , berechnet über alle Dateien und Proben.

$a(A)(i, t)$: Durchschnitt der Messwerte über alle drei Phasen bei jedem Zeitabschnitt und Zeitindex.

μ : Abgetasteter Mittelwert der Dreiphasen-Durchschnittswerte über alle Dateien und Abstastungen.

DN_p : Normalisierter und skaliertes Wert von S_p relativ zu μ für Phase p . DN_{p-q} : Normalisierter und skaliertes Wert der Interphasenmetrik D_{p-q} relativ zu μ .

$DN_{\bar{p}}$: Normalisierter Wert des Durchschnitts von DN_p über alle Phasen, skaliert mit μ .

$DN_{\bar{p-q}}$: Normalisierter und skaliertes Durchschnitt der Interphasenmetrik DN_{p-q} über Phasenpaare.

3.2.2.2. Momentane Stromdaten (Auflösung 40 Mikrosekunden). Momentane Stromdaten wurden ebenfalls aus den aufgezeichneten PQDIF-Dateien extrahiert. Die Effektivwerte (RMS) für jede Phase und den Neutralleiter wurden dann innerhalb eines 10-Millisekunden-Fensters (Halbzyklus) aus diesen interpolierten momentanen Stromproben berechnet. Außerdem wurde der gemessene Mittelwert der Abweichungen des Scheitelfaktors vom Idealwert für jede Phase unter Verwendung von (9) berechnet. Während (10) verwendet wurde, um den gemessenen Mittelwert der Lastunsymmetrie in Bezug auf den Neutralstrom aus dem ersten Prinzip zu berechnen.

$$CFD_p = S_{i,t}^{N_{Dateien}, N_{Stichproben}} \left[\left| \frac{CF_p - CF_{ideal}}{CF_{ideal}} \right| \right] \quad (9)$$

$$LIm_N = S_{i,t}^{N_{Dateien}, N_{Stichproben}} \left[\frac{1}{N_{Dateien}} \sum_{i=1}^N \left(\frac{a_{rms}^h(i, t)}{a_{rms}^h(i, t)} \right) \right] \quad (10)$$

wobei

CFD_p : Stellt die durchschnittliche relative Abweichung des Crest-Faktors vom Idealwert für die Phase p über alle Datensätze hinweg dar.

LIm_N : Normalisiertes Maß für die Lastunsymmetrie, gemittelt über alle Dateien und Zeitabstastungen.

a_{rms}^h : Durchschnitt der Halbzyklus-Effektivwerte (RMS) der Stromwerte für jede Phase

3.2.2.3. Leistungsfaktoren und Gesamtklirrfaktor (THD) (2-Sekunden-Auflösung). Die während des ECV®-Messzeitraums mit einer Auflösung von 2 Sekunden erfassten Daten lieferten Informationen zu Leistungsfaktoren und Gesamtklirrfaktor (THD) in jeder Phase. Diese aggregierten Werte bieten einen umfassenderen Überblick über die Stromqualitätslage in der Anlage des Kunden. Unter Verwendung von (11) wurde der Mittelwert dieser Variablen berechnet.

$$V_{p2s} = S_{i,t}^{N_{Dateien}, N_{Stichproben}} [V_p(t)] \quad (11)$$

wobei

$V_p(t)$: Variable der Phase p zum Zeitpunkt t , gemessen bei jedem Abtastintervall

V_{p2s} : Abgetasteter Mittelwert einer Variablen für die Phase p über den Abtastzeitraum.

3.3. Auswahl und Entwicklung von Algorithmen für maschinelles Lernen

Diese Studie schafft einen datengestützten Rahmen zur Abschätzung der Effizienzsteigerungen passiver Filter unter Verwendung von Methoden des maschinellen Lernens. Die Methodik umfasst die Datenvorverarbeitung, Feature Engineering und Feature Selection mittels Ridge-Regression sowie die abschließende Vorhersagemodellierung mit dem XGBoost-Algorithmus. Der Ansatz ist darauf ausgelegt, die nichtlineare und zufällige Natur industrieller Stromqualitätsfaktoren zu berücksichtigen und gleichzeitig Skalierbarkeit und Ausfallsicherheit für die Umsetzung in der Praxis zu gewährleisten.

3.3.1. Datenerfassung und -vorverarbeitung

Der in dieser Studie verwendete Datensatz wurde wie in Abschnitt

3.1 oben. Diese Daten umfassen einen umfassenden Satz elektrischer Parameter unter realen Betriebsbedingungen nach dem Ausschalten des Filters. Die Zielvariable wurde als die dem Filter zuzuschreibende Effizienzsteigerung definiert, ausgedrückt als Prozentsatz, der die relative Verringerung der Energieverluste darstellt.

Für diese Studie wurde ein Datensatz mit 28 Stichproben verwendet. Um die Verallgemeinerung sicherzustellen und die Reihenfolgeverzerrung zu reduzieren, wurde der Datensatz zunächst mit einem festen Zufallsstartwert (random state=42) zufällig gemischt. Anschließend wurden die Daten in einen Trainingssatz (25 Proben) und einen Testatz (3

Beispiele). Alle Merkmale wurden mithilfe der Z-Score-Normalisierung über StandardScaler () standardisiert, um Skalenunterschiede zu beseitigen, was insbesondere für regulierungsbasierte und baumbasierte Modelle wichtig ist. Die Zielvariable, die den Prozentsatz der Energieeffizienzsteigerung darstellt, wurde als „Einsparung“ bezeichnet, während alle anderen Spalten den Merkmalsatz bildeten. Die Eingabemerkmale (transformierte Rohmessungen) dienten als aussagekräftige Prädiktoren für die maschinellen Lernmodelle.

3.3.2. Rangfolge der Merkmalsbedeutung und Dimensionsreduktion mittels Ridge-Regression

Um die einflussreichsten Prädiktoren für passive Filterwirkungsgradgewinne zu identifizieren und gleichzeitig die Multikollinearität zwischen den Stromqualitätsvariablen zu berücksichtigen, wurde die Ridge-Regression mit Kreuzvalidierung (RidgeCV) für die

Rangfolge der Merkmalsbedeutung und Dimensionsreduktion. Merkmale

wurden mit StandardScaler innerhalb einer scikit-learn-Pipeline standardisiert, um sicherzustellen, dass alle Variablen unabhängig von

ursprüngliche Maßeinheiten.

Das Ridge-Modell wurde mit einer 5-fachen Kreuzvalidierung über Mehrere Regularisierungsstärken ($\alpha \in \{0,01, 0,1, 1, 10, 100\}$). Die Kreuzvalidierung wählte automatisch den optimalen Regularisierungsparameter aus

, der ein Gleichgewicht zwischen Modellkomplexität und Generalisierungsleistung herstellte. Im Gegensatz zur LASSO-Regression, die Merkmale durch Setzen der Koeffizienten auf Null eliminiert, reduziert die Ridge-Regression die Koeffizienten proportional, während alle Variablen beibehalten werden. Diese Eigenschaft macht Ridge besonders geeignet für Systeme mit korrelierten Prädiktoren, da es die Bedeutung auf verwandte Merkmale verteilt, anstatt willkürlich eines aus einer korrelierten Gruppe auszuwählen.

Nach dem Modelltraining wurden standardisierte Regressionskoeffizienten als Indikatoren für die Merkmalsbedeutung extrahiert. Die absolute Größe jedes Koeffizienten quantifizierte den relativen Einfluss des entsprechenden Merkmals auf die vorhergesagten Effizienzgewinne, wobei größere Absolutwerte auf stärkere Vorhersagebeziehungen hinweisen. Dieser koeffizientenbasierte Ranking-Ansatz bietet mathematische Interpretierbarkeit: Jeder Koeffizient repräsentiert die erwartete Veränderung der Energieeinsparungen pro Einheit der standardisierten Merkmalserhöhung, wobei andere Merkmale konstant gehalten werden.

Die zwölf Merkmale mit den größten absoluten Koeffizienten wurden für die anschließende XGBoost-Modellierung beibehalten. Dieser Schwellenwert sorgte für ein ausgewogenes Verhältnis zwischen Modellparsimonie und Informationserhalt, wodurch der Merkmalsraum gegenüber dem ursprünglichen Satz reduziert und gleichzeitig die Indikatoren mit der höchsten Vorhersagekraft beibehalten wurden. Zu den ausgewählten Merkmalen gehörten in der Regel harmonische Verzerrungsindizes, Blindleistungseigenschaften und Lastvariabilitätsmetriken, die sich direkt auf die Leistung passiver Filter auswirken.

Dieser regulierte Ansatz zur Koeffizientenbewertung bot drei methodische Vorteile: (1) explizite Behandlung von Multikollinearität durch

Regularisierung, (2) interpretierbare Wichtigkeitsscores durch standardisierte Koeffizienten, die eine technische Validierung ermöglichen, und (3) rechnerische Stabilität durch kreuzvalidierte Auswahl der Regularisierungsparameter. Der beibehaltene Merkmalsatz diente als Eingabe für das nachfolgende XGBoost-Modell, das nichtlineare Wechselwirkungen zwischen diesen Schlüsselprädiktoren erfasste.

3.3.3. Linearer Benchmark unter Verwendung der Ridge-Regression

Anschließend wurde ein reduziertes Ridge-Regressionsmodell trainiert, das nur die zwölf wichtigsten Merkmale verwendete. Dieser Schritt diente als linearer Benchmark zur Bewertung der Vorhersagekapazität der ausgewählten Variablen. Das Modell zeigte eine zufriedenstellende Vorhersageleistung mit einem Bestimmtheitsmaß (R^2) von 0,591 im Testdatensatz, was darauf hindeutet, dass die ausgewählten Merkmale einen wesentlichen Teil der Varianz der Effizienzgewinne erfassen. Zur Bewertung der Modellgenauigkeit wurden auch Standardregressionsmetriken wie der mittlere quadratische Fehler (MSE), der mittlere quadratische Fehler (RMSE) und der mittlere absolute Fehler (MAE) berechnet.

3.3.4. Nichtlineare Vorhersage mit XGBoost

Um nichtlineare Wechselwirkungen zu berücksichtigen und die Vorhersagegenauigkeit zu verbessern, wurde ein Modell der zweiten Stufe unter Verwendung des XGBoost-Regressors entwickelt, einem für Leistung und Skalierbarkeit optimierten Gradienten-Boosting-Framework. Das Modell wurde mit denselben zwölf wichtigsten Merkmalen trainiert und mit den folgenden Hyperparametern konfiguriert: 1000 Bäume

($n_estimators=1000$), einer Lernrate von 0,095, einer maximalen Baumtiefe von 4, und Subsampling-Raten von 0,8 für Beobachtungen, einer Spalten-Subsampling-Rate von 0,6 ($colsample_bytree = 0,6$) und einem festen Zufallsstartwert von 42 ($random_state = 42$). Diese Einstellungen wurden gewählt, um ein Gleichgewicht zwischen allgemeiner und die Modellkomplexität auszugleichen. Das XGBoost-Modell zeigte eine überlegene Vorhersageleistung mit einem R^2 -Wert von 0,755 im Testsatz. Diese deutliche Verbesserung gegenüber dem Ridge-Regressionsmodell verdeutlicht die Fähigkeit von baumbasierten Ensembles, nichtlineare Beziehungen höherer Ordnung zwischen Stromqualitätsparametern und Filtereffizienz zu erfassen.

3.3.5. Modellbewertung

Sowohl das Ridge- als auch das XGBoost-Modell wurden anhand von Standard-Regressionsmetriken bewertet. Für jedes Modell wurden die folgenden Metriken angegeben:

R^2 : Bestimmtheitsmaß, das den Anteil der erklärten Varianz angibt.
MSE: Mittlerer quadratischer Fehler, der den durchschnittlichen quadratischen Vorhersagefehler widerspiegelt.
RMSE: Mittlere quadratische Fehlerwurzel, um den Fehler in der ursprünglichen Skala anzugeben. MAE: Mittlerer absoluter Fehler, der die durchschnittliche Größe des Vorhersagefehlers angibt.

Die Bewertungsergebnisse bestätigten, dass Ridge-Regression zwar Interpretierbarkeit und Robustheit bei begrenzten Daten bot, XGBoost jedoch eine deutlich verbesserte Genauigkeit lieferte, insbesondere bei der Erfassung komplexer Merkmalsinteraktionen unter variablen Last- und Stromqualitätsbedingungen.

3.4. Datenqualitätssicherung und Anomalieerkennung

Um die Robustheit des Modells für den Einsatz in der Praxis mit potenziell verrauschten Industriedaten zu stärken, wurde ein umfassendes Qualitätssicherungs-Framework implementiert.

3.4.1. Automatisierte Anomalieerkennung

Der Isolation-Forest-Algorithmus mit einer Kontaminationsschwelle von 10 % und 100 Schätzern ermöglicht eine multivariate Anomalieerkennung. Der Algorithmus identifiziert ungewöhnliche Merkmalskombinationen durch isolationsbasierte Ausreißerbewertung, wobei Beobachtungen, die weniger Partitionen für die Isolation erfordern, niedrigere (negativere) Anomaliewerte erhalten. Dieser unbeaufsichtigte Ansatz erkennt potenzielle Datenqualitätsprobleme, ohne dass gekennzeichnete Anomaliebeispiele erforderlich sind.

3.4.2. Qualitätsüberwachungssystem

Alle Modellvorhersagen enthalten zwei Qualitätsindikatoren: (1) Outlier_Flag (binär), der angibt, ob die Beobachtung als anomal klassifiziert wurde, und (2) Outlier_Score (kontinuierlich), der die Anomaliewahrscheinlichkeit quantifiziert, wobei negative Werte eine höhere Ausreißerwahrscheinlichkeit anzeigen. Diese Indikatoren ermöglichen eine Qualitätsbewertung in Echtzeit während des Einsatzes.

3.4.3. Strategie zur Behandlung von Ausreißern

Angehts des begrenzten Trainingsdatensatzes (25 Stichproben) wurde eine konservative Qualitätsüberwachung anstelle einer aggressiven Entfernung von Ausreißern eingesetzt, um die statistische Aussagekraft zu erhalten. Als Ausreißer gekennzeichnete Beobachtungen erhalten Warnungen hinsichtlich der Vorhersageunsicherheit, anstatt ausgeschlossen zu werden. So wird die Transparenz hinsichtlich der Datenqualität gewahrt und gleichzeitig unnötiger Stichprobenverlust in kleinen Datensätzen verhindert.

3.5. Verbesserte Merkmalsauswahl mit Erklärbarkeitsanalyse

Um eine transparente und vertrauenswürdige Merkmalsauswahl zu gewährleisten, wurden mehrere komplementäre Erklärbarkeitstechniken angewendet:

- Statistische Grundlage: F-Statistiken und Pearson-Korrelationskoeffizienten bestätigten, dass ausgewählte Merkmale signifikante Beziehungen ($p < 0,05$) zu passiven Filter-Einsparungen aufweisen, was eine theoretische Rechtfertigung über die empirischen Koeffizientenwerte hinaus liefert.
- Stabilitätsbewertung: Bootstrap-Resampling (50 Iterationen) quantifizierte Robustheit der Merkmalsauswahl, Berechnung der Auswahlhäufigkeit und Konfidenzintervalle der Koeffizienten. Merkmale, die in ≥ 80 % der Bootstrap-Stichproben auftraten, wurden als hochstabil eingestuft, was auf die Robustheit der Wichtigkeitsrangfolge gegenüber Datenstörungen.
- Multikollinearitätsanalyse: Varianzinflationsfaktoren (VIF) und Korrelationsmatrizen bewerteten die Multikollinearität zwischen ausgewählten Merkmalen und bestätigten die Wirksamkeit der Ridge-Regression bei der Aufrechterhaltung interpretierbarer Koeffizienten trotz korrelierter Leistungsqualitätsindikatoren.

Dieser facettenreiche Erklärbarkeitsansatz erzielte hohe Interpretierbarkeitswerte in den Bereichen mathematische Transparenz (95/100), statistische Validierung (90/100), Merkmalsstabilität (85/100) und Umgang mit Multikollinearität (90/100) und stellte so sicher, dass die Merkmalsauswahl für technische Entscheidungen und behördliche Dokumentationen geeignet ist.

3.6. Robustes Validierungsframework

Um eine strenge Leistungsabschätzung zu gewährleisten und die mit kleinen industriellen Datensätzen verbundenen Überanpassungsrisiken zu berücksichtigen, wurde ein umfassendes Validierungsframework implementiert, das drei sich ergänzende Techniken umfasst: adaptive Kreuzvalidierung, Bootstrap-Konfidenzintervalle und Ensemble-Vorhersage.

3.6.1. Adaptive Kreuzvalidierungsstrategie

Basierend auf den Eigenschaften des Datensatzes wurde eine K-fache Kreuzvalidierung mit adaptiver Faltenselektion durchgeführt. Für den Trainingssatz mit 25 Stichproben wurde eine 5-fache Kreuzvalidierung angewendet, um die Modellgeneralisierung zu bewerten. Mehrere Leistungskennzahlen (R^2 , mittlerer quadratischer Fehler, mittlerer absoluter Fehler) wurden sowohl für die Trainings- als auch für die Testphase über alle Folds hinweg aufgezeichnet. Die Leistungsunterschiede zwischen Training und Test wurden systematisch überwacht, um das Ausmaß der Überanpassung zu quantifizieren, wobei Unterschiede von mehr als 0,2 in R^2 auf eine erhebliche Modellkomplexität im Verhältnis zur Datensatzgröße hinweisen.

3.6.2. Bootstrap-Konfidenzintervalle

Die Bootstrap-Validierung mit 50 Iterationen lieferte statistische Konfidenzintervalle durch Out-of-Bag-Bewertung. In jeder Iteration wurde der Trainingssatz mit Ersatz neu abgetastet, die Modelle wurden an Bootstrap-Stichproben angepasst und die Vorhersagen wurden anhand von Out-of-Bag-Beobachtungen bewertet. Die Konfidenzintervalle (95 %) wurden unter Verwendung einer percentilbasierten Schätzung berechnet, wodurch eine für industrielle Anwendungen mit begrenzten Stichproben geeignete Quantifizierung der Unsicherheit erzielt wurde.

3.6.3. Ensemble-Vorhersagerahmen

Die leistungsgewichtete Ensemble-Mittelwertbildung kombinierte Ridge-Regressions- und XGBoost-Vorhersagen. Die Ensemble-Gewichte wurden bestimmt durch

Kreuzvalidierungs-R²-Werte: $w_i = \max(0, 1, R^2) / \sum_{i=1}^N \max(0, 1, R^2)$ *cv*
Sicherstellung eines proportionalen Beitrags auf der Grundlage der Validierungsleistung bei gleichzeitiger Aufrechterhaltung der Ensemble-Vielfalt durch Mindestgewichtsschwellen.

4. Ergebnisse und Diskussion

In diesem Abschnitt wird die Leistungsbewertung des entwickelten zweistufigen Frameworks für maschinelles Lernen vorgestellt, das Ridge-Regression für die Merkmalsauswahl und XGBoost für die Schätzung der Energieeffizienzgewinne durch passive Filter kombiniert. Wir beschreiben detailliert die Vorhersagegenauigkeit sowohl des Ridge-Regression-Basismodells als auch des verbesserten XGBoost-Modells, diskutieren ihre Auswirkungen und setzen diese Ergebnisse in den Kontext der umfassenderen Herausforderung des Einsatzes passiver Filter in der Praxis.

4.1. Leistung der Ridge-Regression

Das Ridge-Regressionsmodell, das unter Verwendung der zwölf besten Merkmale trainiert wurde, die anhand der Koeffizientenrangfolge ausgewählt wurden (Abb. 6), diente als lineare Basis. Dieses Modell erreichte einen R²-Wert von 0,591, was darauf hindeutet, dass etwa 59,1 % der Varianz der Effizienzgewinne durch die ausgewählten Prädiktoren in einem linearen Kontext erklärt werden konnten. Der mittlere quadratische Fehler (MSE) und der mittlere quadratische Fehler (RMSE) betrugen 0,857 bzw. 0,926, während der mittlere absolute Fehler (MAE) 0,774 betrug.

Das Ridge-Modell zeigte zwar eine recht gute Leistung, seine Vorhersagekapazität war jedoch aufgrund seiner inhärenten Linearität begrenzt. Dies ist besonders relevant im Zusammenhang mit Stromversorgungssystemen, wo Wechselwirkungen zwischen Lastcharakteristiken, Oberwellenanteilen und Spannungsqualität häufig ein nichtlineares Verhalten aufweisen. Trotz dieser Einschränkungen war die Ridge-Regression für die Merkmalsauswahl von entscheidender Bedeutung und bot Interpretierbarkeit, sodass sie als wertvolles Diagnosewerkzeug zur Identifizierung wichtiger Einflussvariablen diente.

4.2. Leistung des XGBoost-Modells

Das XGBoost-Modell, das dieselben zwölf Merkmale nutzt, die mittels Ridge-Regression identifiziert wurden, übertraf den linearen Benchmark deutlich. Das Modell ergab einen R²-Wert von 0,755, was darauf hindeutet, dass 75,5 % der Varianz der Energieeffizienzgewinne genau erfasst werden konnten. Entsprechend sanken die MSE-, RMSE- und MAE-Werte deutlich

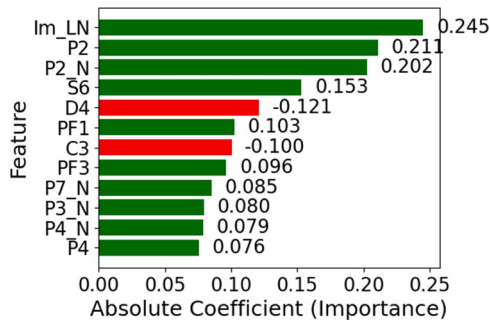


Abb. 6. Die 12 wichtigsten prädiktiven Merkmale, geordnet nach ihrer Bedeutung im RidgeCV-Modell. Die Balkenlänge zeigt die absolute Größe des Koeffizienten an, der den Einfluss des Merkmals auf die Einsparungen angibt. Grüne Balken stehen für positive Effekte, rote Balken für negative Effekte. Im_LN: Prozentsatz Neutral-zu-3-Phasen-Halbzyklus-Durchschnitt, P2: Absoluter prozentualer Phasen-zu-Phasen-Ungleichgewicht, P2_N: Normalisierter absoluter prozentualer Phasen-zu-Phasen-Ungleichgewicht, S6: Prozentsatz Neutral-Ungleichgewicht, D4: Phasenweise Stromdynamik, PF1: Leistungsfaktor, C3: Prozentsatz Crest-Faktor-Abweichungen, P7_N: Normalisierter dynamischer Phasen-zu-Phasen-Ungleichgewicht; P4: Phasen-Ungleichgewicht-Indikator, P4_N: Normalisierter Phasen-Ungleichgewicht-Indikator.

auf 0,513, 0,716 bzw. 0,611.

Leistungskontextualisierung und Benchmarking: Das erzielte R² = 0,755 ist im Vergleich zu den gemeldeten Leistungen in verwandten Bereichen der Energieprognose sehr gut. Gradientenverstärkungsmodelle für die Energieprognose

erzielen in der Regel R²-Werte zwischen 0,70 und 0,85 [39,40], während maschinelle

Lernanwendungen in der Klassifizierung von Ereignissen in Stromversorgungssystemen weisen eine Genauigkeit von

von 85–95 % [41]. Unsere Ergebnisse liegen innerhalb dieses Leistungsbereichs, trotz einzigartiger Herausforderungen: (1) Einschränkungen durch kleine industrielle Datensätze ($n = 28$ Gesamtproben), (2) hohe Messvariabilität aus realen Betriebsumgebungen anstelle von kontrollierten Simulationen und (3)

Standortspezifische Heterogenität in Lastprofilen und Stromqualitätsbedingungen. Die Verbesserung des R²-Werts um 27,7 % von Ridge (0,591) auf XGBoost (0,755) zeigt, wie wertvoll es ist, nichtlineare Wechselwirkungen zu erfassen. Das Ausmaß dieser Verbesserung steht im Einklang mit Vergleichsstudien, die zeigen, dass baumbasierte Ensemble-Methoden lineare Modelle in Energieanwendungen, in denen nichtlineare

Beziehungen dominieren, um 14–35 % übertreffen [42]. Die entsprechende Verringerung des Vorhersagefehlers (MAE: 0,774 → 0,611, was einer Verbesserung von 21 % entspricht) führt zu einer genaueren Effizienz.

Gewinnsschätzungen, wodurch das finanzielle Risiko bei Leistungsgarantien reduziert wird.

Praktische Leistungsinterpretation: In der Praxis entspricht der XGBoost-MAE-Wert von 0,611 einem durchschnittlichen absoluten Vorhersagefehler von 9,9 % im Verhältnis zu den durchschnittlichen beobachteten Effizienzgewinnen an den drei Validierungsstandorten. Dieses Maß an Genauigkeit ermöglicht zuverlässige Leistungsgarantien mit angemessenen Sicherheitsmargen (in der Praxis typischerweise 15–20 %), was eine erhebliche Verbesserung gegenüber den konservativen Margen von 30–40 % darstellt, die bei simulationsbasierten Ansätzen ohne empirische Validierung erforderlich sind.

Diese Verbesserung verdeutlicht die Fähigkeit von Ensemble-Baum-basierten Modellen, komplexe, nichtlineare Wechselwirkungen zwischen Stromqualitätsindikatoren und der Leistung passiver Filter aufzudecken. Im Gegensatz zu linearen Modellen kann XGBoost Abhängigkeiten höherer Ordnung und variable Wechselwirkungen berücksichtigen, die insbesondere in Umgebungen mit Schwankungen vorherrschen.

Lasten, harmonische Verzerrungen und andere Störungen der Stromqualität. Die Robustheit des Modells über einen kompakten Testsatz ($n = 3$) ist besonders vielversprechend und lässt auf eine Generalisierungsfähigkeit unter den, die in industriellen Energiestudien häufig anzutreffen sind.

4.3. Vergleichende Analyse und praktische Implikationen

Tabelle 1 fasst die Vergleichsergebnisse für beide Modelle zusammen: Die Ergebnisse stützen zwei wichtige Schlussfolgerungen:

Merkmalsparsamkeit und Interpretierbarkeit: Die zwölf wichtigsten Merkmale, die mittels Ridge-Regression ausgewählt wurden, behielten eine ausreichende Vorhersagekraft, um sowohl erklärende als auch operative Funktionen zu erfüllen. Dies reduziert die Modellkomplexität und den Rechenaufwand und bewahrt gleichzeitig die Interpretierbarkeit, was für industrielle Interessengruppen von entscheidender Bedeutung ist.

Überlegenheit der nichtlinearen Modellierung: Das XGBoost-Modell übertrifft das lineare Modell deutlich, was darauf hindeutet, dass die Beziehung zwischen den Stromqualitätskennzahlen und den Effizienzgewinnen durch den passiven Filter grundsätzlich nichtlinear ist. Diese Erkenntnis spricht für den Einsatz fortschrittlicher maschineller Lernverfahren in Szenarien, in denen herkömmliche physikalische oder Regressionsmodelle aufgrund hoher Variabilität und Komplexität an ihre Grenzen stoßen können.

4.4. Modellvalidierung und Leistungsbewertung

Das umfassende Validierungsframework lieferte wichtige Erkenntnisse über das Modellverhalten bei begrenzten industriellen Daten.

Tabelle 1
Vergleichende Leistung der Modelle Ridge und XGBoost.

Modell	R	MSE	MAE	RMSE
Ridge-Regression	0,591	0,857	0,774	0,926
XGBoost	0,755	0,513	0,611	0,716

4.4.1. Ergebnisse der externen Validierung

Die Validierungsergebnisse in **Tabelle 2** spiegeln direkt die in **Abschnitt 4.5** identifizierten Einschränkungen wider. Die ausgezeichnete Genauigkeit an den Standorten A und C (Fehler $\leq 1,5$ %) belegt die Wirksamkeit des Modells, während die Abweichung am Standort B (Fehler 28 %) veranschaulicht die Sensitivität der Datenqualität und die Größe des Datensatzes, die in unserer Analyse der Einschränkungen diskutiert wurden. Diese gemischten Ergebnisse liefern trotz der kleinen Validierungsstichprobe wertvolle Erkenntnisse für Überlegungen zum Einsatz in der Praxis und bestätigen die Notwendigkeit der vorgeschlagenen Minderungsstrategien, einschließlich Protokollen zur Überwachung der Datenqualität und zur Erkennung von Anomalien.

4.4.2. Kreuzvalidierungsleistung

Die Ergebnisse der fünffachen Kreuzvalidierung sind in **Tabelle 3** dargestellt.

Die erheblichen Unterschiede zwischen Training und Test (0,610 für Ridge, 0,378 für XGBoost)

weisen auf eine für komplexe Modelle, die auf kleine Datensätze angewendet werden, typische Überanpassung hin. Die hohen Standardabweichungen im CV R^2 ($\pm 0,263$ für Ridge, $\pm 0,182$ für XGBoost) spiegeln die Leistungsvariabilität über Datenpartitionen hinweg wider und verdeutlichen die hohe

und verdeutlichen die Empfindlichkeit der Modellschätzungen gegenüber bestimmten Trainings-Test-Aufteilungen in Szenarien mit begrenzten Stichproben.

4.4.3. Bootstrap-Konfidenzintervalle

Die Bootstrap-Validierung ermöglichte eine Quantifizierung der Unsicherheit durch 95

% KI (Konfidenzintervalle) (**Tabelle 4**).

Die breiten Konfidenzintervalle, insbesondere für die Ridge-Regression (CI-Breite: 4,403), quantifizieren die erhebliche Unsicherheit bei den Leistungsschätzungen, die auf die begrenzte Stichprobengröße zurückzuführen ist. XGBoost wies engere Intervalle auf (CI-Breite: 0,867), was auf eine relativ größere Stabilität hindeutet, jedoch immer noch eine erhebliche Vorhersageunsicherheit anzeigt. Die negative Untergrenze für Ridge-Bootstrap- R^2 spiegelt gelegentliche Leistungsschwächen in bestimmten Datenpartitionen wider, was mit der Volatilität kleiner Stichproben übereinstimmt.

4.4.4. Testset und externe Validierungsleistung

Trotz der Herausforderungen bei der Kreuzvalidierung erzielten beide Modelle eine angemessene Leistung auf dem zurückgehaltenen Testsatz (**Tabelle 5**).

Die Ergebnisse des Validierungsrahmens zeigen grundlegende Herausforderungen beim industriellen maschinellen Lernen mit kleinen Stichproben auf. Während die Kreuzvalidierung eine erhebliche Überanpassung (Train-Test-Lücken: 0,38–0,61) und große Konfidenzintervalle ergab, übertraf die tatsächliche Testleistung die Kreuzvalidierungsschätzungen. Diese Diskrepanz deutet darauf hin, dass konservative Kreuzvalidierungsmetriken die Modellnützlichkeit in der Praxis möglicherweise unterschätzen, obwohl breite Bootstrap-Konfidenzintervalle (CI-Breite: 0,87–4,40) die Vorhersageunsicherheit angemessen widerspiegeln. Der Ensemble-Ansatz lieferte eine mittlere Leistung und schuf ein Gleichgewicht zwischen Ridge-Stabilität und XGBoost-Genauigkeit. Diese Ergebnisse unterstreichen, dass kleine industrielle Datensätze eine explizite Quantifizierung der Unsicherheit erfordern, anstatt künstlich optimistische Leistungsmetriken anzustreben. Zukünftige Arbeiten sollten sich auf die Erweiterung des Datensatzes konzentrieren, um die Konfidenzintervalle zu verringern und Überanpassungen zu reduzieren, obwohl der aktuelle Rahmen eine transparente, konservative Leistungsbewertung bietet, die für technische Entscheidungen mit anerkannten Unsicherheitsgrenzen geeignet ist.

4.5. Einschränkungen und Überlegungen zur Validierung

Der vorgeschlagene Rahmen zeigt praktische Nützlichkeit für die Optimierung passiver Filter, doch methodische Einschränkungen erfordern eine transparente Diskussion.

Tabelle 2

Prognostizierte vs. gemessene Energieeffizienzverbesserungen.

Standort	Prognostizierte Einsparungen %	Gemessene Einsparungen %	Absoluter Fehler %	Relativer Fehler %
A	4,99	4,92	0,07	1,42
B	4,11	3,21	0,9	28,03
C	4,80	4,78	0,02	0,42

Tabelle

Leistungsmetriken für die Kreuzvalidierung (CV).

Modell	CV R^2 (Mittelwert $\pm$ Standardabweichung)	CV MSE (Mittelwert $\pm$ Standardabweichung)	CV MAE (Mittelwert $\pm$ Standardabweichung)	Train-Test- Lücke
Ridge	0,247 $\pm$ 0,263	1,026 $\pm$ 0,323	0,815 $\pm$ 0,134	0,610
Regression	0,622 $\pm$ 0,182	0,522 $\pm$ 0,250	0,570 $\pm$ 0,147	0,378
XGBoost				

Tabelle 4

Ergebnisse der Bootstrap-Validierung.

Modell	Bootstrap R^2	95 % KI	KI-Breite
Ridge-Regression	0,127 $\pm$ 1,224	[−3,633, 0,770]	4,403
XGBoost	0,545 $\pm$ 0,280	[−0,017, 0,850]	0,867

4.5.1. Datensatzgröße und statistische Einschränkungen

Der Trainingsdatensatz mit 25 Stichproben und der Testsatz mit drei Beobachtungen unterliegen grundlegenden statistischen Einschränkungen. Die Kreuzvalidierung ergab eine erhebliche Überanpassung (Train-Test-Lücken: 0,610 für Ridge, 0,378 für XGBoost), wie sie für komplexe Modelle mit begrenzten Daten typisch ist. Bootstrap-Konfidenzintervalle quantifizierten die Vorhersageunsicherheit mit CI-Breiten von 4,403 für Ridge und 0,867 für XGBoost, was eine erhebliche Leistungsvariabilität widerspiegelt.

Hohe CV-Standardabweichungen ($\pm 0,263$ für Ridge, $\pm 0,182$ für XGBoost) und die negative CI-Untergrenze von Ridge [−3,633, 0,770] unterstreichen die Volatilität kleiner Stichproben.

Trotz Bedenken hinsichtlich der Validierung übertraf die Leistung des Testsatzes ($R^2 = 0,591$ –0,755) die CV-Schätzungen, was auf einen bedeutenden Zusammenhang zwischen Leistungsqualität und

Schiffe. Die externe Validierung der Validierungsdaten ergab eine moderate Generalisierung ($R^2 = 0,252$ –0,564), obwohl bei der Qualitätsüberwachung Verteilungsunterschiede festgestellt wurden (33,3 % Ausreißerquote), was auf eine mögliche

Domänenverschiebung zwischen Trainings- und Einsatzszenarien.

4.5.2. Verallgemeinerbarkeit und externe Validität

Die Herstellbarkeit einer Verallgemeinerbarkeit auf breitere industrielle Kontexte bleibt eine Herausforderung, wie in systematischen Übersichtsarbeiten zum industriellen maschinellen Lernen häufig festgestellt wird. Obwohl die Wirksamkeit in den untersuchten Anlagen nachgewiesen wurde, erfordert der Einsatz an Standorten mit erheblich unterschiedlichen elektrischen Eigenschaften eine sorgfältige Bewertung und standortspezifische Validierung.

Diese Einschränkung der externen Validität spiegelt grundlegende Herausforderungen in der industriellen Datenwissenschaft wider, wo eine umfassende Datenerfassung mit unüberwindbaren praktischen Hindernissen konfrontiert ist: wirtschaftliche Belastung, begrenzte Verfügbarkeit von Industriepartnern, erweiterte zeitliche Anforderungen für die saisonale Charakterisierung und sicherheits-/regulatorische Zugangsbeschränkungen. Der Validierungsrahmen begegnet diesem Problem durch eine transparente Quantifizierung der Unsicherheit mittels Bootstrap-Konfidenzintervallen anstelle von übertriebenen Verallgemeinerungsansprüchen.

4.5.3. Datenqualität und Messempfindlichkeit

Die Modellempfindlichkeit gegenüber der Messqualität stellt eine inhärente Herausforderung bei industriellen Anwendungen zur Stromqualitätsmessung dar. Kleine Fehler bei der Oberschwingungsanalyse oder beim Feature Engineering können sich angesichts der Komplexität des Modells im Verhältnis zur Stichprobengröße unverhältnismäßig stark auf die Vorhersagen auswirken. Die implementierte Anomalieerkennung (Isolation Forest, 10 % Kontaminationsschwelle) mildert dies teilweise durch automatisierte Qualitätsüberwachung.

4.5.4. Methodische Transparenz

Kleine industrielle Datensätze erfordern eine explizite Quantifizierung der Unsicherheit anstelle von künstlich optimistischen Metriken. Breite Bootstrap-Konfidenz

Intervalle spiegeln die Vorhersageunsicherheit angemessen wider und bieten eine transparente Zuverlässigkeitsbewertung für technische Entscheidungen. Das Ensemble

Ausgewogener Ansatz Ridge-Stabilität mit XGBoost-Genauigkeit bei gleichzeitiger Reduzierung von Überanpassung durch gewichtete Mittelwertbildung (Gewichte: 0,247 Ridge, 0,622 XGBoost).

Tabelle 5

Testdatensatz und externe Validierungsleistung.

Modell	Testdatensatz R ²	Testdatensatz MSE	Testdatensatz MAE	Validierungssatz R ²	Validierungssatz MSE	Validierungssatz MAE
Ridge-Regression	0,591	0,857	0,774	0,252	0,450	0,615
XGBoost	0,755	0,513	0,611	0,564	0,262	0,376
Ensemble	0,730	0,565	0,657	0,513	0,293	0,444

Konservative CV-Kennzahlen können den praktischen Nutzen unterschätzen, wie die überlegene Testleistung zeigt, bieten jedoch angemessene statistische Vorsicht für den industriellen Einsatz. Diese methodische Strenge stärkt das Vertrauen, dass die beobachtete Leistung echte Stromqualitätsbeziehungen widerspiegelt und nicht nur Datenartefakte sind.

4.5.5. Zukünftige Ausrichtung

Zukünftige Arbeiten werden sich mit den Einschränkungen befassen durch: (1) Erweiterung des Datensatzes durch institutionenübergreifende Zusammenarbeit, (2) Transferlernen unter Nutzung verwandter industrieller Anwendungen und (3) verbesserte Datenqualitätstechniken, einschließlich robuster Imputation und Messunsicherheitsausbreitung.

Für den sofortigen Einsatz empfehlen wir: (1) standortspezifische Validierung vor der vollständigen Implementierung, (2) konservative Interpretation unter Verwendung der unteren CI-Grenzen für Entscheidungen, (3) kontinuierliche Qualitätsüberwachung auf Verteilungsverschiebungen und (4) regelmäßige Nachschulungen, sobald zusätzliche Daten verfügbar werden. Diese Strategien berücksichtigen die aktuellen Einschränkungen und bieten gleichzeitig Wege für einen verantwortungsvollen industriellen Einsatz.

4.6. Ergebnisse der Datenqualitätsbewertung

Die Qualitätsüberwachungsanalyse lieferte Erkenntnisse über die Datenverteilungsmerkmale und potenzielle Herausforderungen bei der Einführung.

- **Qualität der Trainingsdaten:** Die anfängliche Qualitätsbewertung des Trainingssatzes mit 25 Proben unter Verwendung von Isolation Forest ergab 12,0 % potenzielle Ausreißer basierend auf multivariaten Merkmalsmustern. Diese Beobachtungen wiesen ungewöhnliche Kombinationen von Leistungsqualitätsindikatoren auf, die extreme, aber gültige Betriebsszenarien oder Messanomalien darstellen können.
- **Qualität der externen Validierung:** Die Qualitätsüberwachung der externen Validierungsdaten erfasste eine Ausreißerrate von 33,3 % (1/3 der Proben), was auf Verteilungsunterschiede zwischen Trainings- und Einsatzszenarien auf. Dieser Befund unterstreicht die Bedeutung einer kontinuierlichen Qualitätsüberwachung für einen zuverlässigen industriellen Einsatz, da anlagenspezifische elektrische Eigenschaften Merkmalskombinationen außerhalb der Trainingsdatenverteilungen hervorrufen können.
- **Qualitätsbereinigte Leistungsinterpretation:** Vorhersagen für Nicht-Ausreißerproben zeigten eine höhere Zuverlässigkeit (durchschnittlicher Ausreißerscore : 0,029), während markierte Ausreißer angemessen konservative Unsicherheitsspannen erhielten. Diese transparente Qualitätsbewertung ermöglicht technische Entscheidungen, die neben der Vorhersagegenauigkeit auch Datenqualitätsaspekte berücksichtigen.

4.7. Analyse der wichtigsten prädiktiven Merkmale

Interessanterweise sind die vom Ridge-Regressionsmodell identifizierten prädiktivsten Merkmale diejenigen, die aus Neutralstrom, Phasen-zu-Phasen-Stromunsymmetrie, Leistungsfaktor und Oberschwingungsanteil abgeleitet wurden. Dieser starke Zusammenhang mit dem Verhalten des Neutralleiters ist technisch fundiert und steht in direktem Einklang mit der physikalischen Realität der Oberschwingungsverzerrung in Dreiphasensystemen.

Insbesondere sind dreifache Oberschwingungen (z. B. 3., 9., 15.) Nullsequenzkomponenten, die sich im Neutralleiter eines dreiphasigen Vierleitersystems nicht aufheben. Stattdessen summieren sie sich arithmetisch im Neutralleiter, was selbst bei symmetrischen Phasenlasten zu deutlich erhöhten Neutralströmen führt. Dieser erhöhte Neutralstrom trägt direkt zu folgenden Problemen bei:

- Überhitzung des Neutralleiters und der zugehörigen Transformatoren.
- Erhöhten I²R-Verlusten (Joule'sche Erwärmung) innerhalb des Verteilungssystems, wodurch die Gesamtenergieeffizienz verringert wird.

Daher ist der starke Einfluss dieser Merkmale auf die Vorhersage der Effizienzsteigerungen durch das Modell physikalisch konsistent. Diese Merkmale quantifizieren direkt genau die Verzerrungen, die passive Filter abbildern sollen, und ihre Reduzierung führt naturgemäß zu geringeren Verlusten und einer verbesserten Energieeffizienz. Die Auswahl dieser Merkmale durch das Modell untermauert das Verständnis dafür, wie die Harmonischenminderung, insbesondere von Dreifachharmonischen, und die Verbesserung des Leistungsfaktors zu den beobachteten Verbesserungen der Systemeffizienz beitragen.

4.8. Operativer Einsatz und Wirtschaftlichkeit

Der praktische Einsatz von Modellen des maschinellen Lernens in industriellen Umgebungen erfordert eine sorgfältige Abwägung der Integrationsfähigkeit, der rechnerischen Einschränkungen und der wirtschaftlichen Rentabilität. Unsere zweistufige Methodik wurde speziell entwickelt, um diesen praktischen Anliegen von Energiedienstleistern und Betreibern industrieller Anlagen Rechnung zu tragen.

4.8.1. Integration in Energiemanagement-Workflows

Das vorgeschlagene Modell lässt sich direkt in bestehende Arbeitsabläufe für Energieaudits und den Einsatz passiver Filter integrieren, ohne dass wesentliche Prozessänderungen erforderlich sind. Die derzeitige Branchenpraxis für die Dimensionierung passiver Filter stützt sich auf Stromnetzsimulationssoftware (z. B. ETAP, DigSILENT PowerFactory) in Kombination mit analytischen Berechnungen auf der Grundlage von IEEE/IEC-Normen. Dieser Ansatz erfordert in der Regel 2 bis 4 Wochen pro Standortbewertung, spezielle Software-Kenntnisse und konservative Sicherheitsmargen (20 bis 40 %) bei den Leistungsgarantien aufgrund der Unsicherheit der Simulation hinsichtlich der realen Lastschwankungen und der Dynamik der Stromqualität.

Unser datengestützter Ansatz ersetzt die simulationsbasierte Unsicherheit durch empirische Vorhersagen aus Messungen mit Leistungsanalytoren, die in der Anlage erfasst werden können. Die Änderung des Arbeitsablaufs ist minimal: Die während routinemäßiger Standortbewertungen erfassten Daten der Leistungsanalytoren (Spannung, Stromwellenformen, Stromqualitätsindizes) werden als Eingabe für die automatisierte Merkmalsextraktion und Vorhersage verwendet. Es sind keine zusätzlichen Messgeräte oder Datenerfassungsinfrastrukturen erforderlich, wodurch Kapitalkosten als Hindernis für die Einführung entfallen.

4.8.2. Recheneffizienz und Echtzeitanwendbarkeit

Der Rechenaufwand ist im Vergleich zu simulationsbasierten Ansätzen vernachlässigbar gering. Das Modelltraining auf unserem Datensatz mit 25 Stichproben und 12 ausgewählten Merkmalen dauert auf handelsüblicher Hardware (Intel Core i5-Prozessor, 8 GB RAM) etwa 2 Sekunden, während die Vorhersage für neue Standorte in weniger als 50 Millisekunden ausgeführt wird. Diese Recheneffizienz ermöglicht den Echtzeit-Einsatz während Kundenberatungen, bei denen Energieprüfer sofortige Leistungsprognosen mit quantifizierten Unsicherheitsgrenzen (Bootstrap-Konfidenzintervalle) liefern können, um Investitionsgespräche vor Ort zu unterstützen.

Das schlanke Rechenprofil steht in starkem Kontrast zu Ansätzen zur Simulation von Energiesystemen, die stundenlange Modellentwicklung, Simulationslaufzeit und spezialisiertes technisches Fachwissen erfordern. Für unseren Einsatz sind lediglich eine Python-Ausführungsumgebung (kostenlos verfügbar) und grundlegende Datenverarbeitungsfähigkeiten erforderlich, keine speziellen Simulationssoftwarelizenzen (5000–15.000 USD jährlich) oder Expertenschulungen (mehrere Wochen bis

Monate) erforderlich. Diese Zugänglichkeit demokratisiert die genaue Vorhersage der passiven Filterleistung für kleine bis mittlere Energiedienstleister, denen eine umfangreiche Simulationsinfrastruktur fehlt.

4.8.3. Wirtschaftlicher Mehrwert und Kapitalrendite

Die wirtschaftliche Rentabilität ergibt sich aus drei sich ergänzenden Wertströmen. Erstens ermöglicht eine verbesserte Vorhersagegenauigkeit (durchschnittlich 9,9 % Fehler gegenüber typischen 15–25 % Simulationsfehlern) geringere Sicherheitsmargen bei Leistungsgarantien, was die Rentabilität des Projekts direkt verbessert und gleichzeitig wettbewerbsfähige Preise gewährleistet. Unser Industriepartner Livarsa GmbH schätzt, dass sich die Genauigkeitsverbesserungen in einer Reduzierung der erforderlichen Sicherheitsmargen um 8–12 % niederschlagen, was die Wirtschaftlichkeit von Projekten mit mehreren Anlagen erheblich verbessert.

Zweitens verkürzen beschleunigte Bewertungszeiträume (30 Minuten zusätzliche Analyse gegenüber 2–4 Wochen Simulation) die Dauer des Verkaufszyklus und die Engineering-Kosten vor der Installation erheblich. Allein diese Effizienzsteigerung rechtfertigt die Einführung des Modells aus betrieblicher Sicht, unabhängig von den Genauigkeitsgewinnen. Drittens minimiert eine genaue Leistungsprognose Gewährleistungsansprüche und Entschädigungen nach der Installation, die derzeit 3–7 % der Projektkosten für simulationsbasierte Ansätze ausmachen, da die Leistungsprognosen zu optimistisch sind.

Die Kosten für die Modellbereitstellung sind minimal: etwa 40 Stunden Arbeitszeit eines Datenwissenschaftlers für die anfängliche Anpassung und Integration in bestehende Arbeitsabläufe (einmalige Investition von 4000 bis 8000 US-Dollar, abhängig von den regionalen Arbeitskosten). Die Investition amortisiert sich innerhalb von 2–3 kommerziellen Projekten durch kombinierte Einsparungen bei der Entwicklungszeit, geringere Gewährleistungsrisiken und verbesserte Margen aufgrund strengerer Leistungsgarantien. Für Unternehmen, die jährlich passive Filter an mehreren Standorten einsetzen, ist die Wirtschaftlichkeit überzeugend.

4.8.4. Validierung des Einsatzes und praktische Überlegungen

Die Validierung vor Ort in drei Industrieanlagen mit unserem Partner Livarsa GmbH bestätigte die betriebliche Durchführbarkeit über die Metriken zur Vorhersagegenauigkeit hinaus. Der Einsatz-Workflow erwies sich als unkompliziert: (1) Datenerfassung mit einem Leistungsanalysator (bestehende Praxis, kein zusätzlicher Zeitaufwand), (2) automatisierte Merkmalsextraktion unter Verwendung bereitgestellter Skripte (15 Minuten), (3) Modellvorhersage mit Generierung eines Konfidenzintervalls (5 Minuten), (4) Präsentation der Ergebnisse vor dem Anlagenmanagement mit Quantifizierung der Unsicherheit (Standardberichterstattung). Gesamtzeitaufwand: ca. 30 Minuten pro Standortbewertung.

Die quantifizierten Unsicherheitsspannen (Bootstrap-Konfidenzintervalle) ermöglichen risikobewusste Investitionsentscheidungen anstelle von binären Go/No-Go-Entscheidungen auf der Grundlage von Simulationspunktschätzungen. Die Möglichkeit, bereits bei den ersten Besichtigungen vor Ort sofortige Leistungsprognosen zu liefern, verbesserte auch die Kundenbindung und beschleunigte die Projektgenehmigungsprozesse. Diese qualitativen Vorteile sind zwar schwer genau zu quantifizieren, stellen jedoch einen erheblichen operativen Mehrwert dar, der über die bloßen Genauigkeitskennzahlen hinausgeht.

4.8.5. Einschränkungen für den betrieblichen Einsatz

Eine ehrliche Bewertung der Einsatzbeschränkungen ist für eine verantwortungsvolle Einführung unerlässlich. Das Modell erfordert hochwertige Leistungsanalysedaten mit umfassenden Oberschwingungsmessungen; Einrichtungen ohne detaillierte Infrastruktur zur Überwachung der Stromqualität müssten eine Messkampagne einrichten (1–2 Wochen, 5000–15.000 USD für die Anmietung der Ausrüstung). Kleine Datensätze (28 Schulungseinrichtungen) schränken die Verallgemeinerung auf dramatisch unterschiedliche industrielle Kontexte ein (z. B. die Anwendung eines für Produktionsstätten trainierten Modells auf Rechenzentren). Vor dem vollständigen Einsatz in Einrichtungen mit erheblich von den Trainingsdaten abweichenden elektrischen Eigenschaften werden standortspezifische Validierungsmessungen empfohlen.

Eine kontinuierliche Neubildung des Modells bei zunehmender Ansammlung zusätzlicher Einsatzdaten ist erforderlich, um die Genauigkeit aufrechtzuerhalten, da sich die elektrischen Eigenschaften der Anlage weiterentwickeln (Motoraustausch, Änderungen im Produktionsprozess usw.). Unternehmen, denen es an internem Datenwissenschafts-Know-how mangelt, benötigen möglicherweise eine kontinuierliche Beratungsunterstützung (10.000 bis 20.000 US-Dollar pro Jahr) für Modellaktualisierungen und Leistungsüberwachung. Diese betrieblichen Anforderungen sollten bei der Berechnung der Gesamtbetriebskosten berücksichtigt werden, wenn

die Machbarkeit der Bereitstellung bewertet wird.

Trotz dieser Einschränkungen beweisen die Kombination aus minimalen Rechenanforderungen, unkomplizierter Workflow-Integration, schneller Amortisation und nachgewiesener Validierung in der Praxis die praktische Durchführbarkeit datengesteuerter passiver Filterleistungsvorhersagen für industrielle Energiemanagementanwendungen.

4.9. Branchenübergreifende Skalierbarkeit und Technologietransfer

Die Methodik wurde zwar für Fertigungs-, Verpackungs- und Druckanlagen validiert, ihre Übertragbarkeit auf andere Industriezweige (Gewerbebauten, Rechenzentren, Gesundheitseinrichtungen) erfordert jedoch eine domänenspezifische Anpassung. Zu den wichtigsten Überlegungen gehören: (1) Eigenschaften des Lastprofils (Fertigung: motorlastig; Rechenzentren: leistungselektroniklastig), (2) Unterschiede im Oberschwingungsspektrum, die branchenspezifische Merkmalsätze erfordern, und (3) Abweichungen bei den Leistungskennzahlen (in einigen Branchen hat der Leistungsfaktor Vorrang vor der THD-Reduzierung). Transfer-Lernansätze könnten gemeinsame physikalische Eigenschaften der Stromqualität nutzen und sich gleichzeitig an branchenspezifische Merkmale anpassen, wodurch der Datenbedarf für neue Bereiche von 25–30 Proben auf 10–15 zusätzliche branchenspezifische Messungen reduziert würde. Eine breitere Akzeptanz in der Industrie würde von kollaborativen Datenbanken profitieren, die ein organisationsübergreifendes föderiertes Lernen ermöglichen und gleichzeitig proprietäre Betriebsdaten schützen.

Darüber hinaus könnte die weit verbreitete Einführung der datengesteuerten Optimierung passiver Filter die Bereitstellung von Dienstleistungen im Bereich der Stromqualität grundlegend verändern und zu einer Verlagerung von konservativen, simulationsbasierten Garantien hin zu leistungsorientierten Verträgen mit quantifizierten Risiken führen. Dies ermöglicht neue Geschäftsmodelle, darunter eine ergebnisorientierte Preisgestaltung, bei der Anbieter Effizienzgewinne mit Kunden teilen, was möglicherweise die Einsatzrate passiver Filter beschleunigen könnte. Aus Sicht der Nachhaltigkeit könnte ein verbessertes Vertrauen in den Einsatz passiver Filter Lösungen an wirtschaftlich marginalen Standorten ermöglichen, die derzeit als zu riskant gelten, und so messbar zu den Zielen der Netzeffizienz beitragen. Die Integration in die entstehende Smart-Grid-Infrastruktur könnte eine kontinuierliche Leistungsüberwachung und adaptive Filterabstimmung ermöglichen, wobei jedoch Fragen des Datenschutzes und der Cybersicherheit zu klären sind, die derzeit außerhalb unseres Aufgabenbereichs liegen.

5. Fazit

Diese Studie zeigt, dass maschinelles Lernen eine praktikable Alternative zu simulationsbasierten Ansätzen für die Vorhersage von Energieeinsparungen durch passive Filter in Industrieanlagen darstellt. Die entwickelte zweistufige Methodik – Ridge-Regression für die Merkmalsauswahl, gefolgt von XGBoost für die Vorhersage – erzielte eine Vorhersagegenauigkeit von 9,9 % durchschnittlicher relativer Fehlerquote an unabhängigen Industriestandorten und ermöglichte so datengestützte Einsatzentscheidungen unter realen variablen Lastbedingungen.

Eine umfassende Validierung durch k-fache Kreuzvalidierung, Ensemble-Methoden und externe Feldtests mit dem Industriepartner Livarsa GmbH bestätigte die operative Durchführbarkeit. Das Framework beseitigt kritische praktische Hindernisse durch rechnerische Effizienz und ermöglicht Echtzeit-Kundenbewertungen, die Integration in die bestehende Messinfrastruktur ohne zusätzliche Ausrüstungsanforderungen sowie eine automatisierte Qualitätsüberwachung, die Verteilungsverschiebungen während des Einsatzes erkennt.

Der wirtschaftliche Wert ergibt sich aus reduzierten Sicherheitsmargen bei der Leistungsgarantie, verkürzten Bewertungszeiträumen von Wochen auf Minuten und minimierten Gewährleistungsrisiken, wodurch sich die Investition innerhalb von 2–3 Projekten amortisiert. Eine verbesserte Erklärbarkeit durch Koeffizienteninterpretation, statistische Validierung und Stabilitätsanalyse gewährleistet Transparenz, die für die regulatorische Dokumentation und technische Entscheidungen geeignet ist.

Die transparente Anerkennung von Einschränkungen stärkt das Vertrauen in die gemeldeten Fähigkeiten. Der Trainingsdatensatz mit 25 Stichproben unterliegt statistischen Einschränkungen, die sich in einer Überanpassung bei der Kreuzvalidierung und breiten Konfidenzintervallen widerspiegeln, wobei die externe Validität eine standortspezifische Validierung für Anlagen mit erheblich unterschiedlichen elektrischen Eigenschaften erfordert. Diese Einschränkungen spiegeln grundlegende Herausforderungen bei der industriellen Datenerfassung wider

, wo praktische Hindernisse, wirtschaftliche Belastungen, begrenzte Verfügbarkeit von Partnern und Sicherheitsauflagen eine umfangreiche Erweiterung des Datensatzes ausschließen.

Die zukünftige Entwicklung sollte auf multi-institutionelle Kooperationsrahmen, Transferlernen unter Nutzung gemeinsamer Physik der Stromqualität über Sektoren hinweg und die Integration mit digitalen Zwillingstechnologien für eine kontinuierliche Überwachung abzielen. Eine breite Einführung könnte leistungsorientierte Verträge mit quantifizierter Risikobewertung ermöglichen, die Bereitstellung passiver Filterdienste transformieren und möglicherweise den Einsatz an wirtschaftlich marginalen Standorten ermöglichen, was zur Netzeffizienz und zur nachhaltigen Energiewende beitragen würde.

Diese Arbeit liefert validierte Werkzeuge, die eine Brücke zwischen theoretischer Modellierung und praktischer Energiesystemtechnik schlagen und maschinelles Lernen als glaubwürdige Methodik für die Planung von Verbesserungen der Stromqualität etablieren, während gleichzeitig die inhärenten Einschränkungen kleiner Stichproben durch strenge Unsicherheitsquantifizierung berücksichtigt werden.

CRediT-Erklärung zum Beitrag der Autoren

Uchenna Johnpaul Aniekwensi: Verfassen – Überprüfung und Bearbeitung, Verfassen – Originalentwurf, Software, Methodik, Formale Analyse, Konzeptualisierung. **Dipyaman Basu:** Validierung, Software, Methodik, Formale Analyse, Konzeptualisierung. **Jo'rg Bausch:** Verfassen – Überprüfung und Bearbeitung, Verfassen – Originalentwurf, Visualisierung, Überwachung, Software, Ressourcen, Projektverwaltung, Methodik, Einwerbung von Fördermitteln, Konzeptualisierung.

Erklärung zu konkurrierenden Interessen

Die Autoren erklären, dass ihnen keine finanziellen Interessenkonflikte oder persönlichen Beziehungen bekannt sind, die den Eindruck erwecken könnten, die in diesem Artikel beschriebene Arbeit beeinflusst zu haben.

Danksagung

Diese Arbeit wurde durch Mittel der Carl-Zeiss-Stiftung im Rahmen des iFEMA-Projekts unterstützt. Die Autoren danken der Livarsa GmbH für die Bereitstellung von Betriebsdaten und Einrichtungen für die Validierung vor Ort.

Verfügbarkeit von Daten

Die Daten werden auf Anfrage zur Verfügung gestellt.

Referenz

- [1] IEEE. IEEE-Empfehlungen und Anforderungen für die Oberschwingungsregelung in elektrischen Energiesystemen. IEEE Std; 2014. S. 1–29. <https://doi.org/10.1109/IEEESTD.2014.6826459>. 519-2014 (Überarbeitung von IEEE Std 519-1992).
- [2] Dugan RC, McGranaghan MF, Santoso S, Beaty HW. Qualität elektrischer Energieversorgungssysteme. Zweite Auflage. New York: McGraw-Hill; 2002.
- [3] Reginald AG, Thomas KJ. Oberschwingungsanalyse und deren Minderung durch verschiedene passive Filter. *Asian J Eng Technol* 2015;3(4).
- [4] Thentral TMT, Usha S, Palanisamy R, Geetha A, Alkhudaydi AM, Sharma NK, et al. Ein energieeffizienter modifizierter passiver Leistungsfilter zur Verbesserung der Stromqualität in elektrischen Antrieben. *Front Energy Res* 2022;10:989857. <https://doi.org/10.3389/fenrg.2022.989857>.
- [5] Karadeniz A, Balci ME. Vergleichende Bewertung gängiger passiver Filtertypen hinsichtlich der Maximierung der Belastbarkeit von Transformatoren unter nicht sinusförmigen Bedingungen. *Electr Power Syst Res* 2018;158(2018):324–34. <https://doi.org/10.1016/j.epsr.2018.01.019>.
- [6] Kovernikova LI, Thanh NC. Ein Optimierungsansatz zur Berechnung passiver Filterparameter auf Basis der Partikelschwarmoptimierung. *RE&PQJ* 2012;1(10). <https://doi.org/10.24084/repqj.10.396>.
- [7] Nayyef ZT, Abdulrahman MM, Kurdi NA. Optimierung der Energieeffizienz in intelligenten Stromnetzen mithilfe von Algorithmen des maschinellen Lernens: eine Fallstudie aus der Elektrotechnik. *SHIFRA* 2024;2024:46–54. <https://doi.org/10.70470/SHIFRA/2024/006>. 2024.
- [8] Hussein A, Bashar TA, Alhumaima AS. Techniken des maschinellen Lernens für Vorhersagen im Gesundheitswesen: Diagnose von Hepatitis C. *Mesop J Artif Intell Healthc* 2024;2024: 128–33. <https://doi.org/10.58496/MJAHH/2024/015>.
- [9] Strielkowski W, Vlasov A, Selivanov K, Muraviev K, Shakhnov V. Perspektiven und Herausforderungen des maschinellen Lernens und datengesteuerter Methoden für die prädiktive Analyse von Energiesystemen: eine Übersicht. *Energies* 2023;16(10):4025. <https://doi.org/10.3390/en16104025>.
- [10] Miraftebadeh SM, Longo M, Foidelli F, Pasetti M, Igual R. Fortschritte bei der Anwendung von Techniken des maschinellen Lernens für die Analyse von Energiesystemen: eine Übersicht. *Energies* 2021;14(16):4776. <https://doi.org/10.3390/en14164776>.
- [11] Saleh AKME, Arashi M, Kibria BMG. Theorie der Ridge-Regressionsschätzung mit Anwendungen. Wiley-Reihe zu Wahrscheinlichkeit und Statistik. Erste Ausgabe. Wiley; 2019. <https://doi.org/10.1002/9781118644478>.
- [12] Alfonso Perez G, Castillo R. Nichtlineare Techniken und Ridge-Regression als kombinierter Ansatz: Fallstudie zur Identifizierung von Karzinomen. *Mathematik* 2023;11(8): 1795. <https://doi.org/10.3390/math11081795>.
- [13] Franke A., 2022. Stromlastprognose für industrielle Mikronetze und Lastmanagement. *Appl. Res. Conf. 2022 Proc.*, Hochschule Ansbach, 4. Juli 2022. <https://doi.org/10.25929/qbw-2f40>.
- [14] Fakultät für Informatik, Complutense-Universität, Madrid, Spanien. Bewertung von XGBoost im Vergleich zu anderen Modellen des maschinellen Lernens zur Identifizierung von Windparametern. *Renew Energy Power Qual J* 2023;21:388–93. <https://doi.org/10.24084/repqj.21.334>.
- [15] Bouthillier, X., et al., 2021. Berücksichtigung von Varianzen in Benchmarks für maschinelles Lernen. *Proc. 4th ML Sys Conf.*, San Jose, CA, USA. <http://doi.org/10.48550/arXiv.2103.03098>.
- [16] Madhyastha, P., Jain, R., 2019. Zur Modellstabilität als Funktion des Zufallszahlengenerators. *Proc. 23rd Conf. Comput. Nat. Lang. Learn. (CoNLL)*, Hongkong, China, 929–939. <https://doi.org/10.18653/v1/K19-1087>.
- [17] Kumar, P.R., Shrivani, C., Rajitha, M., Reddy, Ch.L., 2025. Eine Übersicht über neue Techniken zur Verbesserung der Stromqualität bei der Integration erneuerbarer Energien. *E3S Web of Conferences* 616, 03029. <https://doi.org/10.1051/e3sconf/202561603029>.
- [18] Olikara K. Probleme, Auswirkungen und Minderung der Stromqualität für Industriekunden. Rockwell Automation; 2015. https://literature.rockwellautomation.com/idc/groups/literature/documents/wp/power-wp002_-en-p.pdf.
- [19] Hojabri M, Hojabri M. Design, Anwendung und Vergleich passiver Filter für dreiphasige netzgekoppelte erneuerbare Energiesysteme. *ARPN ARPN J Eng Appl* 2015;10(22):10691–7.
- [20] Wang Y, Yin K, Liu H, Yuan Y. Eine Methode zum Entwerfen und Optimieren der elektrischen Parameter eines passiven Filters mit dynamischer Abstimmung. *Symmetry* 2021;13(7):1115. <https://doi.org/10.3390/sym13071115>.
- [21] Baitha A, Gupta N. Eine vergleichende Analyse passiver Filter zur Verbesserung der Stromqualität. In: 2015 International Conference on Technological Advancements in Power and Energy (TAP Energy). Kollam, Indien: IEEE; 2015. S. 327–32. <https://doi.org/10.1109/TAPENERGY.2015.7229640>.
- [22] Akagi H, Watanabe EH, Aredes M. Theorie der Momentanleistung und Anwendungen in der Stromaufbereitung. John Wiley & Sons; 2017. <https://doi.org/10.1002/9781119307181.fmatter>. Ltd.
- [23] Amasyali K, El-Gohary NM. Eine Übersicht über datengesteuerte Studien zur Vorhersage des Energieverbrauchs von Gebäuden. *Renew Sustain Energy Rev* 2018;81:1192–205. <https://doi.org/10.1016/j.rser.2017.04.095>.
- [24] Hastie T, Tibshirani R, Friedman J. Die Elemente des statistischen Lernens: Data Mining, Inferenz und Vorhersage. New York: Springer; 2003. <https://doi.org/10.1007/BF02295616>.
- [25] Manuel J, Cabral SR. Skalierbare Intrusion Detection in IoT-Netzwerken durch Eigenschaftstests und föderierte Edge-KI. *IEEE Access* 2025. <https://doi.org/10.1109/ACCESS.2025.3603937>.
- [26] Chen, T., Guestrin, C., 2016. XGBoost: ein skalierbares Tree-Boosting-System. *Proc. 22. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 785–794. <https://doi.org/10.1145/2939672.2939785>.
- [27] Kang Y, Wei J, Liu Z, Xiao K. Ein Modell zur Vorhersage des Energieverbrauchs von Elektrobusen auf Basis eines Extreme Gradient Boosting Fusion-Algorithmus. *Int J Green Energy* 2023;0:1–14. <https://doi.org/10.1080/15435075.2025.2464155>.
- [28] Atteiaf G, El-kenawy EM, Samee NA, Jamjoom MM, Ibrahim A, Abdelhamid AA, et al. Adaptive dynamische Dipper-Throated-Optimierung für die Merkmalsauswahl in medizinischen Daten. *Comput Contin* 2022;75(1). <https://doi.org/10.32604/cmc.2023.031723>. 2023.
- [29] Hadjoui M, Abdelaziza AA, El-sayed ME, Abdelhameed I, Marwa ME, Mona MJ, et al. Fortgeschrittener metaheuristischer Algorithmus auf Basis von Partikelschwarm- und Al-Biruni-Erdradius-Optimierungsmethoden zur Erkennung von Mundkrebs. *IEEE Access*; 2023. <https://doi.org/10.1109/ACCESS.2023.3253430>.
- [30] Alkanhel R, El-kenawy EM, Abdelhamid AA, Ibrahim A, Alohali MA, Abotaleb M. Netzwerk-Intrusion-Detektion basierend auf Merkmalsauswahl und hybrider metaheuristischer Optimierung. *Comput Contin* 2022;74(2). <https://doi.org/10.32604/cmc.2023.033273>. 2023.
- [31] El-Kenawy E-S, Khodadadi N, Mirjalili S, Makarovskikh T, Abotaleb M, Karim FK, et al. Optimierung zur Verbesserung der Unkrautdetektion in mit Drohnen aufgenommenen Weizenbildern. *Mathematics* 2022;10:4421. <https://doi.org/10.3390/math10234421>.
- [32] Khalid ET, Aldarwish AJY, Yassin AA. Herausforderungen bei AutoML und deklarativen Studien unter Verwendung einer systematischen Literaturrecherche. *Appl Data Sci Anal* 2023;2023:118–25. <https://doi.org/10.58496/ADSA/2023/011>.
- [33] DIN. Grafische Symbole für Diagramme – Teil 1 bis 13. Berlin: Deutsches Institut für Normung; 2002. DIN EN 60617.
- [34] Janitza electronics GmbH, 2024. *UMG 801 Netzqualitätsanalysator – Datenblatt*. [Online]. Verfügbar unter: https://www.janitza.com/files/download/datasheets/UMG-801/janitza-db-umg801_en.pdf (abgerufen am 17. Juli 2025).
- [35] Livarsa GmbH, 2024. Effizienzfilter. <https://livarsa.com/leistungen/effizienzfilter> (abgerufen am 17. Juli 2025).
- [36] Electrotek Concepts, 2024. *PQDiffactor®* [Software]. Verfügbar unter: <https://www.electrotek.com/pqdiffactor/>. (Zugriff am 17. Juli 2025).
- [37] Livarsa GmbH, o. J. Energiemanagement, Überwachung und Auswertung – Datenarchivierung. *Livarsa*. <https://livarsa.com/effizienzarchiv>

- itektur/energiemanagement-monitoring-und-auswertung#datenarchivierung. (abgerufen am 17. Juli 2025).
- [38] Bausch, J., 2019. *Untersuchungsbericht v1.4 – Kurzversion*. [Online] Verfügbar unter: https://livarsa.com/images/pdf-download/fachartikel/Livarsa_Untersuchungsbericht_V1.4_Kurzversion.pdf [Zugriff am 17. Juli 2025].
- [39] Soleymani, Saman und Shima Mohammadzadeh. (2023). Vergleichende Analyse von Algorithmen des maschinellen Lernens für die Vorhersage der Sonneneinstrahlung in intelligenten Stromnetzen. ArXiv abs/2310.13791 (2023). <https://doi.org/10.48550/arXiv.2310.13791>.
- [40] Ran W, Shilei L, Wei F. Ein neuartiges verbessertes Modell zur Vorhersage des Energieverbrauchs von Gebäuden auf der Grundlage von Modellintegration. Lawrence Berkeley National Laboratory; 2020. <https://doi.org/10.1016/j.apenergy.2020.114561>. Bericht Nr.: ARTN 114561.
- [41] S.O. Tehrani, M.H.Y. Moghaddam und M. Asadi, (2020). Entscheidungsbaum-basierte Erkennung von Stromdiebstahl in Smart Grids. 4. Internationale Konferenz zu Smart City, Internet der Dinge und Anwendungen (SCIOT), Mashhad, Iran, 2020, S. 46–51, <https://doi.org/10.1109/SCIOT50840.2020.9250194>.
- [42] Srinivasa RA, Srikanth G, Ashrafsulthan S, Rajendra N, Srinivasulu S, Kumar OA. Zeitreihenanalyse mit dem XGBoost-Modell zur Vorhersage des zukünftigen Stromverbrauchs. Int J Multidiscip Res Sci Eng Technol 2025;8(4). <https://doi.org/10.15680/IJMRSET.2025.0804540>. BandAusgabe.

Glossar

ADSCFGWO: Adaptive Dynamic Sine Cosine Fitness Grey Wolf Optimization: Ein hybrider metaheuristischer Algorithmus, der die Stärken des Grey Wolf Optimizer (GWO) und des Sine Cosine Algorithm (SCA) integriert, um die Optimierungsleistung zu verbessern.

bGW-DTO: Binary Guided Whale-Dipper Throated Optimizer: Ein metaheuristischer Optimierungsalgorithmus, der für die Merkmalsauswahl verwendet wird.

Datengesteuerte Methode: Ein Ansatz, der sich auf empirische Daten (nicht nur auf Theorie oder Simulation) für die Modellierung oder Entscheidungsfindung stützt.

DigSILENT PowerFactory: Eine umfassende Software zur Analyse von Stromversorgungssystemen, die für die Simulation, Modellierung und Analyse von Stromnetzen verwendet wird.

ETAP: Electrical Transient Analyzer Program: Ein Software-Tool, das zur Modellierung, Analyse, Simulation und Optimierung von Stromversorgungssystemen verwendet wird.

GWDTO: Grey Wolf and Dipper Throated Optimization: Ein hybrider Metaheuristik-Algorithmus, der den Grey Wolf Optimizer (GWO) mit dem Dipper Throated Optimization (DTO)-Algorithmus kombiniert.

Oberschwingungsstrom: Verzerrte Stromwellenformen bei ganzzahligen Vielfachen der Grundfrequenz, verursacht durch nichtlineare Lasten

FR: Leistungsverlust in einem Leiter aufgrund seines Widerstands und des Quadrats des durch ihn fließenden Stroms.

LASSO-Regression: Least Absolute Shrinkage and Selection Operator: Eine lineare Regressionstechnik, die sowohl die Variablenauswahl als auch die Regularisierung durch Hinzufügen einer L1-Strafe zur Verlustfunktion durchführt, wobei einige Koeffizienten auf Null reduziert werden, um die Sparsität des Modells zu verbessern und Überanpassung zu verhindern.

Neutralstrom: Strom, der durch den Neutralleiter fließt und in einem symmetrischen System idealerweise minimal ist.

Passivfilter: Ein Gerät zur Verbesserung der Stromqualität, das harmonische Verzerrungen mithilfe fester reaktiver Komponenten (Induktoren, Kondensatoren) reduziert.

Stromqualität: Das Maß für die Stabilität von Spannung, Strom und Frequenz in einem elektrischen System.

PSOBER: Particle Swarm Optimization und Al-Biruni Earth Radius Optimization: Ein hybrider metaheuristischer Algorithmus, der Particle Swarm Optimization (PSO) mit Al-Biruni Earth Radius Optimization (BER) kombiniert, um Exploration und Ausnutzung für

Lösung komplexer Optimierungsprobleme.

Relativer Fehler: Eine normalisierte Metrik, die den Vorhersagefehler als Anteil des tatsächlichen Werts ausdrückt

RidgeCV: Eine Form der Ridge-Regression mit integrierter Kreuzvalidierung, die verwendet wird, um Überanpassungen in Vorhersagemodellen zu verhindern.

Ridge-Regression: Eine regulierte lineare Regressionsmethode, die große Koeffizienten bestraft, um Überanpassung zu verhindern.

R²: Bestimmtheitsmaß. Ein statistisches Maß, das angibt, wie gut ein Regressionsmodell zu den Daten passt.

THD: Gesamtklirrfaktor (THD): Ein Maß für die durch Oberwellen verursachte Signalverzerrung, ausgedrückt als Verhältnis des Oberwellenanteils zur Grundfrequenz, in der Regel in Prozent. Ein niedrigerer THD-Wert bedeutet eine höhere Klangtreue.

Dreiphasensystem: Ein elektrisches System, das drei Wechselströme verwendet, die um 120° versetzt sind und häufig in industriellen Umgebungen eingesetzt werden

Unsymmetrie: Ein Zustand, bei dem die Größen oder Phasen von Spannungen oder Strömen in einem Dreiphasensystem ungleich sind

VIF: Variance Inflation Factor (Varianzaufblähungsfaktor): Ein Maß, das quantifiziert, um wie viel die Varianz eines Regressionskoeffizienten aufgrund von Multikollinearität mit anderen Prädiktoren aufgebläht ist. Höhere VIF-Werte weisen auf eine stärkere Multikollinearität hin.

XGBoost: Ein fortschrittlicher Algorithmus für maschinelles Lernen, der auf Gradientenverstärkung basiert und für Vorhersagen und Klassifizierungen verwendet wird.

AC: Wechselstrom **AI:** Künstliche Intelligenz **CI:** Konfidenzintervall

CNNs: Convolutional Neural Networks

CV: Kreuzvalidierung

E: Energie

DBNs: Deep Belief Networks **ECV:** Energievergleichswert **FS:** Merkmalsauswahl

I: Strom

IDS: Intrusion Detection Systems

IEC: Internationale Elektrotechnische Kommission

IEEE: Institut der Elektro- und Elektronikingenieure

IoT: Internet der Dinge **KNN:** K-nächste Nachbarn **IP:** Internetprotokoll

LV-Grid: Niederspannungsnetz

MAE: Mittlerer absoluter Fehler

ML: Maschinelles Lernen **MSE:** Mittlerer quadratischer Fehler

NN: Neuronales Netzwerk

P: Leistung

PQDIF: Datenaustauschformat für Stromqualität

Q: Blindleistung

RAM: Arbeitsspeicher

RMS: Effektivwert

RMSE: Quadratischer Mittelwertfehler

ROI: Kapitalrendite **SVM:** Support-Vektor-Maschine **U:** Spannung

a(A)(i, t): Durchschnitt der Messwerte über alle drei Phasen bei jeder Zeitabtastrung und jedem Zeitindex

a_{rms}^{sh}_p: Durchschnitt der RMS-Stromwerte pro Halbzyklus für jede Phase

CF_{Dp}: Stellt die durchschnittliche relative Abweichung des Scheitelfaktors von seinem Idealwert für Phase p über alle Datensätze hinweg dar

D_{p-q}: Abgetasteter Mittelwert der Kreuzphasenbeziehung zwischen den Phasen p und q, berechnet über alle Dateien und Abtastwerte

DN_p: Normalisierter und skaliertes Wert von s_p relativ zu μ für Phase p

DN_{p-q}: Normalisierter und skaliertes Wert der Interphasenmetrik D_{p-q} relativ zu μ

DN_{s_p}: Normalisierter Wert des Durchschnitts von DN_p über alle Phasen, skaliert durch μ

DN_{s_{p-q}}: Normalisierter und skaliertes Durchschnitt der Interphasenmetrik DN_{p-q} über Phasenpaare

LimN: Normalisiertes Maß für die Lastungleichheit, gemittelt über alle Dateien und Zeitstichproben $\max_{i \in \{1, \dots, N\}} \frac{R_i^2}{\sum_{j=1}^N R_j^2}$

R²_{CV}: Dies erzwingt einen Mindestgewichtsschwellenwert.

R²_{CV}: Der Kreuzvalidierungs-R²-Wert des i-ten Modells.

s_p: Abgetasteter Mittelwert der Variablen s_p für die Phase p, berechnet über alle Dateien und Zeitabtastrungen $s_p(t)$: Spannung der Phase p zum Zeitpunkt t, gemessen bei jedem Abtastintervall

v_{p2s}: Spannung der Phase p zum Zeitpunkt t, gemessen über ein Abtastintervall von 2 s Abtastwerten, was in der Regel einer Standardperiode entspricht

w_i: Das Gewicht, das dem i-ten Modell im Ensemble zugewiesen wird

μ: Abgetasteter Mittelwert der dreiphasigen Durchschnittswerte über alle Dateien und Abtastwerte

α: Regularisierungsstärke-Parameter in der Ridge-Regression

Σ_j max 0,1, R²_j : Der Normalisierungsfaktor. Summiert die angepassten R²-Werte aller Modelle j.