



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

REGIONALE MILCHMENGENPROGNOSE: REGRESSIONSMODELLE UND MASCHINELLES LERNEN IM VERGLEICH

Dominik Baaken und Sebastian Hess

dominik.baaken@uni-hohenheim.de

Institut für Agrarpolitik und Landwirtschaftliche Marktlehre,
Universität Hohenheim, Schwerzstr. 46, 70599 Stuttgart



2021

Vortrag anlässlich der 61. Jahrestagung der GEWISOLA
(Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.)

***„Transformationsprozesse im Agrar- und Ernährungssystem:
Herausforderungen für die Wirtschafts- und Sozialwissenschaften,
22. bis 24. September 2021***

REGIONALE MILCHMENGENPROGNOSE: REGRESSIONSMODELLE UND MASCHINELLES LERNEN IM VERGLEICH

Dominik Baaken¹ und Sebastian Hess

Zusammenfassung

Prognose-Instrumente für das regionale Milchaufkommen werden seit dem Wegfall der europäischen Milchquoten im Jahr 2015 seitens der Marktteilnehmer vermehrt nachgefragt, aber existieren bisher nicht. Bestehende Ansätze zur Milchmengenprognose basieren meist auf der Vorhersage der Laktationskurve einzelner Milchkühe. In der vorliegenden Arbeit werden sechs Modellansätze aus verschiedenen Bereichen des Maschinellen Lernens (ML) und der linearen Regression (OLS) miteinander verglichen. Für die Ansätze werden unterschiedliche Variablenblöcke miteinander kombiniert, um zeitliche Trends, direkte und indirekte Wettereffekte, sowie das Preisgeschehen in die Vorhersage mit einzubeziehen. Für einen Vorhersagezeitraum von 21 Monaten kann für die einbezogenen landwirtschaftlichen Betriebe aus Niedersachsen ein maximales Bestimmtheitsmaß von 0,92 durch ML-Methoden und maximal 0,77 durch OLS Regressionen erzielt werden. Im Vergleich der ML-Algorithmen untereinander tritt ein Unterschied der Modelle vor allem im Hinblick auf die Trainingsgeschwindigkeiten zu Tage.

Keywords

Milchmarkt, Milchproduktion, Vorhersagemodelle, Maschinelles Lernen.

1 Einleitung

Die deutsche Milchproduktion ist mengenmäßig die Viertgrößte der Welt (FAOSTAT, 2020). Innerhalb der letzten 10 Jahre ist die produzierte Milchmenge im Bundesgebiet um etwa 11,7 % angestiegen (BLE, 2020). Gleichzeitig ist der Milchpreis in diesem Zeitraum sehr volatil mit einem Schwankungsbereich von 20,6 bis 42,4 ct/l (BMEL, 2020). Seit dem Ende der Milchquote im europäischen Raum im Jahr 2015 können Landwirte ihre Produktionskapazitäten deutlich einfacher ausweiten. Eine Folge dessen ist ein Informationsdefizit unter Marktteilnehmern über die produzierte Milchmenge der kommenden Monate.

Einige Molkereien versuchen die anfallende Milchmenge auf Basis freiwilliger Selbstauskünfte und Selbsteinschätzung der Landwirte abzuschätzen. Für den Aufbau eines Prognoseindex zur Vorhersage der Milchlieferungsmenge genügt es jedoch nicht, Landwirte lediglich zu ihrer Produktionsabsicht zu befragen, da nicht alle produktionsrelevanten Einflüsse von allen Betrieben gleichermaßen jederzeit korrekt eingeschätzt werden können (LARCHER et al., 2015). Aus ökonomischer Sicht ist bei Selbstauskünften zudem mit strategischen Antworten zu rechnen. Die Herausforderung für die Prognose von Milchmengen ist durch die Abhängigkeit von langsamen, biologischen Prozessen bedingt (HOEHL und HESS, 2021). Diese erschweren Landwirten die direkte und feinjustierte Anpassung ihrer Produktionsmenge. Seit vielen Jahren werden ökonometrische Modelle auf Grundlage der Laktationskurve einzelner Milchkühe und dadurch für einen gesamten Betrieb geschätzt (ADEDIRAN et al., 2012). Vorhersagemodelle für eine größere Anzahl an Betrieben oder gar ganze Regionen existieren bisher jedoch nicht oder sind nicht öffentlich zugänglich. Dabei könnten Methoden des Maschinellen Lernens (ML) eine

¹ Institut für Agrarpolitik und Landwirtschaftliche Marktlehre, Universität Hohenheim, 70599 Stuttgart, dominik.baaken@uni-hohenheim.de

Möglichkeit bieten, eine praxisrelevante Prognose auf Basis sehr großer Datensätze zu gewährleisten. Anstelle der Bildung und Anwendung eines Modells auf Basis bestehender Theorie, werden bei ML Muster aus vorhandenen Datenstrukturen abgelesen, aus denen Schlüsse auf zukünftige Entwicklungen gezogen werden können (MULLAINATHAN und SPIESS, 2017). Der Fokus von ML-Methoden liegt dabei insbesondere auf der Vorhersagegenauigkeit während sich ökonometrische Ansätze eher auf die statistischen Eigenschaften von Parametern konzentrieren (STORM et al., 2020).

In dieser Arbeit werden unterschiedliche Prognoseansätze des ML auf die Vorhersage von Milchmengen angewendet und beispielhaft verglichen. Als Referenz dient dabei ein herkömmliches Regressionsmodell. So soll die Frage geklärt werden, welche Art von ML-Algorithmen für die Vorhersage regionaler Milchproduktion angewendet werden kann und wie sie sich gegenseitig unterscheiden. Zudem ist von Interesse, welche Kombination und Konfiguration von Trend-, Wetter- und Preisvariablen für die Prognose von Relevanz sind. Dazu wird zuerst die Datenstruktur und -vorbereitung beschrieben und anschließend eine Einführung in die verwendeten Modelle und Methoden gegeben. In Abschnitt 4 werden Ergebnisse der Algorithmen dargestellt und diskutiert. Im Anschluss werden Schlussfolgerungen für zukünftige Prognosemodelle gezogen.

2 Datengrundlage

Es wird unterstellt, dass die Milchmenge eines landwirtschaftlichen Betriebes in Niedersachsen eine Funktion des Tierbestandes, der Herdenleistung, der Marktpreise, des Managements und weiterer Faktoren ist. Diese Einflussgrößen stehen jedoch in einer komplexen innerbetrieblichen Wechselwirkung zueinander und bedingen sich in ihrer Wirkung auf den Milchoutput zum Teil gegenseitig. Aus Platzgründen wird hier daher auf einen konzeptionellen Rahmen wie zum Beispiel eine Distanzfunktion verzichtet.

Der verwendete Datensatz bezieht sich auf monatliche Beobachtungen im Zeitraum 9/2010 bis 9/2019 und fällt damit in etwa in den zeitlichen Rahmen, in dem die Abschaffung der Milchquote bereits beschlossen war und durch eine jährliche Erhöhung der Quotenmengen vorbereitet wurde („weiche Landung“). Variablen der einzelbetrieblichen Milchproduktion sowie Wetter- und Preisdaten werden in die Modelle mit einbezogen.

Die Daten zur Milchproduktion werden vom Landeskontrollverband Niedersachsen e.V. (LKV) zur Verfügung gestellt. Der Verband erstellt monatlich eine Milchleistungsprüfung für alle Mitgliedsbetriebe aus Niedersachsen und stellt diese Daten für 49 zufällig ausgewählte Milchviehbetriebe bereit. Auf diese Weise entstehen Datenpunkte für ca. 7.500 Milchkühe aus Niedersachsen, welche durch einen Mittelwert für die jeweiligen Betriebe aggregiert werden. Somit ergibt sich ein einzelner Monatswert als Mittelwert der gemolkene Kühe pro Betrieb. Dieser Wert wird mit 30,5 Tagen und der Anzahl gemolkener Tiere pro Betrieb und Monat multipliziert, sodass dieser der durchschnittlich produzierten Menge Milch je Betrieb und Monat entspricht. Betrachtet wird dabei der „energy corrected milk yield“ (ECM), berechnet nach SJAUNJA et al. (1990). Die Anzahl der gehaltenen Tiere pro Betrieb variiert von unter 30 bis über 500 Tiere. Der Mittelwert liegt bei ca. 150 Kühen pro Betrieb; der Großteil der Betriebe (ca. 60 %) befindet sich in der Größenklasse der Betriebe mit 50-199 Milchkühen. Dies entspricht etwa der Verteilung der Größenklassen in Niedersachsen. Betriebe mit über 200 Kühen sind dagegen im Vergleich mit fast 30 % Anteil überrepräsentiert (in Niedersachsen im Mai 2020: knapp 10 %) (LSN, 2020). Dies ist aber im Hinblick auf eine Prognose eher als hilfreich zu bewerten, da der Strukturwandel weiter voranschreitet und sich der Anteil der Größenklassen fortlaufend in Richtung der größeren Betriebe verschiebt (BLE, 2020).

Es muss davon ausgegangen werden, dass die Milchproduktion in einer Region in erheblichem Maße durch kurz- und langfristige klimatische Einflüsse bestimmt wird (HOEHL und HESS, 2021). Dies berührt insbesondere die Qualität des Grundfutters und die Physiologie der Tiere.

Es erscheint daher notwendig, auch regionale Wetter- und Klimadaten in die Prognose mit einzubeziehen. Für die Wettervariablen kann auf die „COSMO Regional Reanalysis“-Datenbank des Deutschen Wetterdienstes (DWD) zurückgegriffen werden. In diesen numerischen Wettermodellen werden Beobachtungen und Analysen miteinander kombiniert, sodass eine Vielzahl an Parametern für ein Gitter von ca. 6 km x 6 km Auflösung zur Verfügung steht.

Die verwendeten Preise für Milch werden vom Bundesministerium für Landwirtschaft und Ernährung (BMEL) bereitgestellt und beziehen sich auf konventionell erzeugte Kuhmilch in Niedersachsen bei tatsächlichem Fett- und Eiweißgehalt. Für Futterpreise wird der durch das Statistische Bundesamt ermittelte Index für Einkaufspreise landwirtschaftlicher Betriebe verwendet.

3 Methoden

In diesem Abschnitt werden die verschiedenen Modelle zur Vorhersage der Milchproduktion in Niedersachsen vorgestellt. Das Basismodell ist eine multiple lineare Regression auf Basis des Kleinst-Quadrat-Schätzers (OLS). Im Vergleich zu diesem Basismodell werden eine Reihe von Maschinellen Lernmodellen (ML) aufgestellt, um einerseits den Unterschied zum ökonometrischen Ansatz und andererseits die verschiedenen ML-Algorithmen für die Vorhersage der Milchproduktion zu vergleichen.

Die betrachteten Betriebe lassen sich anhand ihrer jährlichen Wachstumsrate der erzeugten Milchmenge in zwei Klassen einteilen. Zur Berücksichtigung der verschiedenen Arten landwirtschaftlicher Betriebe wird eine Indikatorfunktion eingeführt, bevor die Algorithmen auf den Datensatz angewendet werden. Durch diese indizierte Unterscheidung können die Algorithmen die Muster in den Daten besser erlernen und für neue Daten verallgemeinern. Die Indikatorfunktion ist wie folgt definiert:

$$(1) \quad I_{WR}(x) = \begin{cases} 1 & \text{wenn } x_i \leq WR \\ 0 & \text{wenn } x_i > WR \end{cases}$$

x_i ist die durchschnittliche, jährliche Wachstumsrate der Milchproduktion in Betrieb i und $WR = Q_{0,8}(x)$. Durch diese Aufteilung anhand des 0,8-Quantils werden für alle Modellspezifikationen k jeweils ein Untermodell für Gruppe a mit $x_i > WR$ und für Gruppe b mit $x_i \leq WR$ gebildet. Für die Betrachtung im Ergebnisteil werden beide Gruppen wieder je Modellspezifikation k zu einer Gesamtmenge zusammengefügt.

Abb. 1a: Milchproduktion im Zeitverlauf

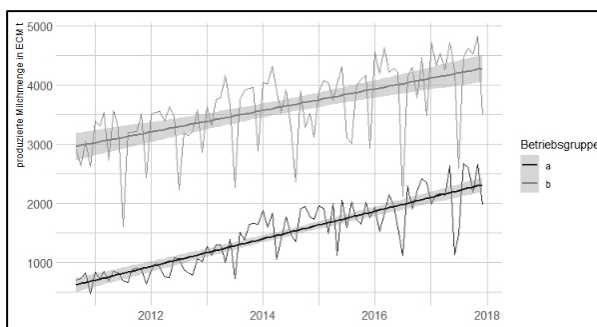
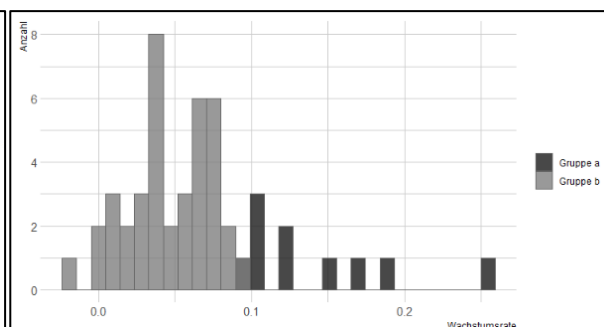


Abb. 1b: Histogramm Wachstumsraten



Quelle: Eigene Darstellungen.

Der Großteil der Betriebe (80 %) weist eine jährliche Änderung der Milchmenge zwischen ca. -1,5 % und etwa 9 % auf. Die übrigen Betriebe wachsen dagegen mit jährlichen Wachstumsraten von durchschnittlich ca. 14 % und erkennbaren Sprüngen in der monatlich produzierten Menge (siehe Abbildung 1). Dies lässt darauf schließen, dass diese Betriebe nicht nur durch betriebsinterne Remontierung, sondern durch größere Investitionen auf Basis zugekaufter Tiere wachsen. Die Betriebe können daher in zwei Gruppen eingeteilt werden, die

sich anhand ihrer Wachstumsrate unterscheiden. Die daraus resultierende, verschiedene Entwicklung der produzierten Milchmenge im Zeitverlauf ist in Abbildung 1a dargestellt. Die zu erkennende Trendlinie zeigt, dass in Gruppe a – stark wachsende Betriebe – die Produktionsmenge schneller ansteigt als in Gruppe b – moderat wachsende Betriebe.

Zur weiteren Verarbeitung wird der Datensatz der jeweiligen Gruppen a und b in ein Trainings- und ein Test-Set unterteilt. Auf diese Weise können die Modelle und Algorithmen die Muster des Datensatzes im Trainingsdatensatz erlernen. In einem zweiten Schritt kann die Präzision der Algorithmen in der Vorhersage durch Anwendung auf das Test-Set ermittelt werden. Dabei wird eine typische Aufteilung in etwa 80 % Trainingsdaten und rund 20 % Testdaten verwendet. Um eine Vorhersage für die kurz- und mittelfristige Milchmengenproduktion über einen Zeitraum von 21 Monaten abzubilden, werden hierfür die aktuellsten Daten von 1/2018 bis 9/2019 als Test-Set verwendet.

3.1 Ausgangsmodell: Prognose der monatlichen Milchproduktionsmenge

Das Ausgangsmodell unterstellt, dass die Milchproduktion eines Betriebes i durch ein Zusammenwirken ökonomischer und klimatischer Gegebenheiten im Monat t und vorherige Monate (Lags) erklärt werden kann. Ökonomische Effekte lassen sich in Preiseffekte für Inputs und Outputs sowie betriebsindividuelle Managementfaktoren unterscheiden. Der Einfluss des Wetters auf die Milchleistung kann in direkte und indirekte Effekte unterschieden werden. Direkte Wettereinflüsse sind typischerweise Hitzestress für Milchkühe und die Beeinträchtigung ihres Stoffwechsels (HILL und WALL, 2017). Indirekte Wettereinflüsse ergeben sich durch Qualitäts- und Quantitätsunterschiede der Fütterung (HARDER et al., 2019). Das Ausgangsmodell kann wie folgt beschrieben werden:

$$(2) \quad y_{it} = T_t + W_{it} + P_t + \mu_i + \varepsilon_{it}$$

mit y_{it} = durchschnittlich monatlich produzierte Menge Milch [kg ECM] von Betrieb i zum Zeitpunkt (Monat) t ; T_t = Zeitvariablen; W_{it} = Wettervariablen; P_t = Preisvariablen; μ_i = Dummy für Betrieb i ; ε_{it} = Fehlerterm. Die Zeitvariablen sind in Datums-, Jahres- und Faktorvariablen für jeden Monat aufgeteilt. Der Abschnitt der Wettervariablen enthält zehn verschiedene Wetteraspekte zum Zeitpunkt t , welche den Koordinaten des Betriebes i zugeordnet werden können. Diese lauten wie folgt: Niederschlag [mm], minimale Temperatur [°C], maximale Temperatur [°C], mittlere Temperatur [°C], Bodentemperatur [°C], relative Luftfeuchtigkeit [%], Sonnenscheindauer [h], Wolken-Bedeckungsgrad [%], Windgeschwindigkeit [km/h], Temperatur-Feuchtigkeits-Index (berechnet nach MADER et al. (2006)). Als Preisvariablen werden die monatlichen Milchpreise für Niedersachsen [ct/l] und ein Futterpreisindex je zum Zeitpunkt t verwendet.

Tabelle 1 bildet für jedes Modell die Kombination der Variablengruppen ab. Es werden unterschiedliche Modellspezifikationen gebildet, um den Einfluss der Variablengruppen auf die Vorhersagegenauigkeit zu vergleichen.

Tabelle 1: Übersicht Modellspezifikationen

Gleichung	Modell	Wetter	Interaktionsterm Wetter und Monatsdummy (MD)	Wetter zeitverzögert (Lags)	Interaktionsterm Wetter Lags und Monatsdummy (MD)	Preise	Dummy je Betrieb
		W_{it}	$W_{it} * MD_m$	$W_{i;t-n}$	$W_{i;t-n} * MD_{m-n}$	P_t	μ_i
2	A	X				X	X
3	B		X			X	X
4	C				X	X	X
5	D		X		X	X	X
6	E	X	X	X	X	X	X

Quelle: Eigene Darstellung.

Das Ausgangsmodell bildet nur die direkten Wettereffekte ab. Für die weiteren Modelle wird ein Interaktionsterm zwischen den direkten Wettervariablen und einem Dummy MD_m für jeden Monat m eingeführt mit $m = \{Januar, \dots, Dezember\}$. Dieser Interaktionsterm wird in Gleichung (3) in die Modellspezifikation B mit einbezogen:

$$(3) \quad y_{it} = T_t + W_{it} * MD_m + P_t + \mu_i + \varepsilon_{it}$$

Zur Modellierung indirekter Wettereffekte wird das Modell (3) um zeitverzögerte Lags der Wetter-Variablen ergänzt, während der Term der direkten Effekte nicht berücksichtigt wird. Dies führt zu der folgenden Gleichung:

$$(4) \quad y_{it} = T_t + W_{i,t-n} * MD_{m-n} + P_t + \mu_i + \varepsilon_{it}$$

mit W_{t-n} = Wettervariablen zum Zeitpunkt $t - n$ und MD_{m-n} = Dummy für Monat $m - n$, wobei $n = \{1, \dots, 6\}$. Alle weiteren Variablen sind wie im Ausgangsmodell definiert.

Diese beiden Modelle der direkten und indirekten Effekte werden zusätzlich in einem Modell als Kombination beider Effekte zusammengenommen, sodass das Modell wie folgt aufgebaut ist:

$$(5) \quad y_{it} = T_t + W_{it} * MD_m + W_{i,t-n} * MD_{m-n} + P_t + \mu_i + \varepsilon_{it}$$

Als letztes Modell werden alle verschiedenen Modelltypen zu einem Modell für alle erstellten Variablen kombiniert, das wie folgt dargestellt wird:

$$(6) \quad y_{it} = T_t + W_{it} + W_{it} * MD_m + W_{i,t-n} + W_{i,t-n} * MD_{m-n} + P_t + \mu_i + \varepsilon_{it}$$

Eine Einbindung von Lags der Preisvariablen kann entfallen, da Preistrends und Preiserwartungen einerseits in der betriebspezifischen Wachstumsrate und andererseits in der Trendvariable T enthalten sind.

3.2 Modelle des Maschinellen Lernens

Die im vorangegangenen Abschnitt vorgestellten Modellspezifikationen werden durch unterschiedliche Schätzmethoden auf den Datensatz angewendet, wobei herkömmliche Regressionen² (OLS) mit sog. „überwachten Lernmethoden“ des Maschinellen Lernens (ML) verglichen werden. Bei ML werden die Variablen als Features bezeichnet. Es gibt eine Vielzahl von überwachten ML-Ansätzen, daher beschränkt sich die Analyse auf drei verschiedene Modelltypen: regularisierende Regression, baumbasierte Methoden und einen Ansatz eines neuronalen Netzes.

3.2.1 Regularisierende Regression

Regularisierung basiert auf der Ergänzung einer linearen Regression um einen Strafterm, der die Koeffizienten gegen Null schrumpfen lässt, wodurch Koeffizienten mit geringem Erklärungsgehalt „bestraft“ und somit aus dem Modell gedrängt werden (JAMES et al., 2013). Dieses Verfahren ist besonders effektiv für Datensätze mit einer großen Anzahl erklärender Variablen, da nur Koeffizienten, die von Null abweichen und eine wesentliche Vorhersagekraft beitragen, auch in das Modell aufgenommen werden (STORM et al., 2020). Die bekanntesten Regularisierungsverfahren sind *LASSO* und *Ridge Regression*. Deren Optimierungsfunktionen werden durch die Minimierung der Summe der Residuenquadrate (RSS) in (7) und einen zusätzlichen Strafterm wie in Gleichung 8 nach JAMES et al. (2013) dargestellt. Dabei werden die Koeffizienten $\hat{\beta}_j$ minimiert.

$$(7) \quad RSS = \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$$

² Es wird kein Fixed Effects-Modell geschätzt, da dieses zwar das Modell um einen Individualeffekt der Betriebe korrigieren könnte, dies aber hier nicht notwendig ist, da nicht die Kausalität, sondern die Vorhersagegenauigkeit primär untersucht werden soll.

$$(8) \quad \text{LASSO: } \min_{\beta_j} (RSS + \lambda \sum_{j=1}^p |\beta_j|); \text{ Ridge: } \min_{\beta_j} (RSS + \lambda \sum_{j=1}^p \beta_j^2)$$

Beide Regularisierungsmethoden verwenden einen Tuning-Parameter λ , der durch die Daten mittels Kreuzvalidierung bestimmt werden kann. Der Einfluss der Regularisierung wächst mit zunehmendem λ (JAMES et al., 2013). Die Methoden unterscheiden sich durch die Konfiguration ihrer Strafterme, die auch L_1 - (LASSO) und L_2 -Term (Ridge) genannt werden. Der Strafterm des elastischen Netzes (EN) ist in Gleichung 9 dargestellt und bildet eine Kombination aus beiden Straftermen mit $\alpha \in [0,1]$ (ZOU und HASTIE, 2005).

$$(9) \quad (1 - \alpha)\lambda \sum_j |\beta_j| + \alpha\lambda \sum_j \beta_j^2$$

Im vorliegenden Fall wird $\alpha = 0,5$ angenommen und der Tuning-Parameter λ mit einer 10-fachen Kreuzvalidierung validiert.

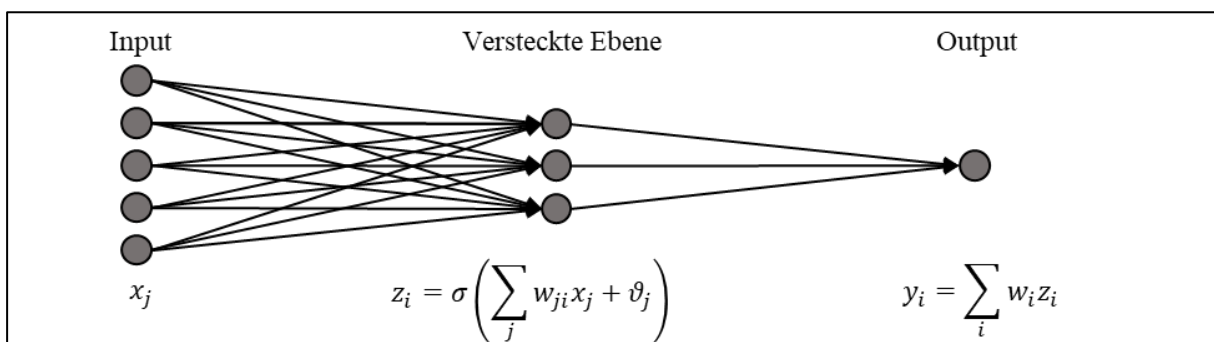
3.2.2 Entscheidungsbäume

Entscheidungsbäume dienen der Einteilung eines Datensatzes in maximal homogene Untergruppen anhand von bestehenden Features (STORM et al., 2020). So entstehen durch einen sogenannten „Classification and Regression Tree“ (CART) jeweils binäre Splits (BREIMANN et al., 1984). Für den Algorithmus „Random Forest“ wird der Datensatz zufällig in Untergruppen von Beobachtungen und Features aufgeteilt, die sich jeweils nur leicht unterscheiden. Für jede Untergruppe entsteht ein Entscheidungsbaum, aus dem der Algorithmus einen Durchschnitt und daraus eine Vorhersage bildet (BREIMANN, 2001). Eine weitere Möglichkeit die Vorhersagekraft von Entscheidungsbäumen zu optimieren, ist die Methode des Boostings. Dabei werden die Residuen der vorherigen Modelle verwendet, sodass der nachfolgende Baum sich auf diese spezialisieren kann (SCHAPIRE, 1990). Für die Vorhersage der Milchproduktion in Niedersachsen werden zwei Boosting Algorithmen angewendet: Xtreme Gradient Boosting (XGB) und Gradient Boosting Machine (GBM). Für beide werden jeweils folgende Hyperparameter ausgewählt: Anzahl Bäume = 2000; Lernrate = 0,01; Maximale Tiefe = 5, Anteil Subgruppe = 0,65; 10-fache Kreuzvalidierung.

3.2.3 Neuronales Netz

Typische neuronale Netze (NNs) oder Multilayer Perceptrons (MLPs) sind aufgebaut wie in Abbildung 2 dargestellt. Sie bestehen aus mehreren verbundenen Schichten mit jeweils zahlreichen Neuronen. Ihr Ziel ist es, eine Funktion f zu approximieren, die aus den Ausgangswerten (Inputs) x lernt, um einen Ausgangsvektor (Output) y zu definieren. NNs sind im Grunde eine Abbildung von $y = f(x; w)$, wobei die Gewichte w kontinuierlich aktualisiert werden, bis sie die beste Approximation der Funktion ergeben (GOODFELLOW et al., 2016). Die Gewichte und Neuronen in der Mitte des Netzes können nicht direkt beobachtet werden, daher werden sie als versteckte Ebene bezeichnet (HASTIE et al., 2017).

Abbildung 2: Schematischer Aufbau eines neuronalen Netzes



Quelle: Eigene Darstellung nach HASTIE et al. (2017).

Die Stärke der Verbindung zwischen den Neuronen wird durch die Gewichte dargestellt, die mit den Input-Werten multipliziert und zu einer Konstante ϑ addiert werden, die als Startwert für die Optimierung fungiert (ATHEY und IMBENS, 2019). Die Summe dieser Produkte wird über die Aktivierungsfunktion σ der Art „rectified linear units“ (ReLU) an die versteckte Schicht gegeben (siehe Gleichung 10 und Abbildung 2).

$$(10) \quad \sigma(z) = \max\{0, z\}$$

Im speziellen Fall von Abbildung 2 werden die aktivierten Werte dann in die Ausgabeschicht eingespeist, wobei dies nur als Vereinfachung dient. NNs werden aufgrund ihrer Verknüpfung verschiedener Funktionen als Netze bezeichnet. In diesem Fall werden drei versteckte Ebenen verwendet, sodass eine Kettenstruktur in Form von $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$ auftritt. Die Aktualisierung und Optimierung der Gewichte erfolgt durch eine Verlustfunktion, ähnlich wie bei der Optimierung einer linearen Regression (GOODFELLOW et al., 2016). Für die Anwendung des NNs zur Milchmengenprognose werden die folgenden Hyperparameter gewählt: 3 versteckte Ebenen mit 200, 100 und 50 Neuronen; Optimierung mit Backpropagation; 150 Epochen; Stopp-Kriterium = RMSE; minimale Batch-Größe = 32; 10-fache Kreuzvalidierung.

3.2.4 Stacking

Die Methode des Stacking kombiniert dem Namen entsprechend mehrere Algorithmen über einen sogenannten „Super-Lerner“ miteinander. Hierbei soll die bestmögliche Kombination von unterschiedlichen Modellen die Vorhersage verbessern. Der Lern-Algorithmus kann beliebig gewählt werden, wobei die Anwendung des Super-Lerners voraussetzt, dass die vorherigen Modelle jeweils mit derselben V -fachen Kreuzvalidierung trainiert wurden. Die dabei entstandenen vorhergesagten Werte der Basis-Lerner L werden nun in eine Matrix mit der Dimension $V \times L$ kombiniert, die als Datengrundlage für den Super-Lerner dient (VANDER LAAN et al., 2007). Aus diesem Grund wird die Methode des Stackings auch als eine verfeinerte Version der Kreuzvalidierung bezeichnet, die das „Winner-takes-it-all“-Prinzip der Kreuzvalidierung verbessert (WOLPERT, 1992). In diesem Fall wird ein Random-Forest-Algorithmus zur Kombination des EN, GBM und NN angewendet. Das XGB-Modell wird hierbei nicht in das Stacking-Modell mit einbezogen, da die Lernweise dem des GBM sehr ähnelt und daher nur wenig Mehrwert entstehen würde.

4 Ergebnisse und Diskussion

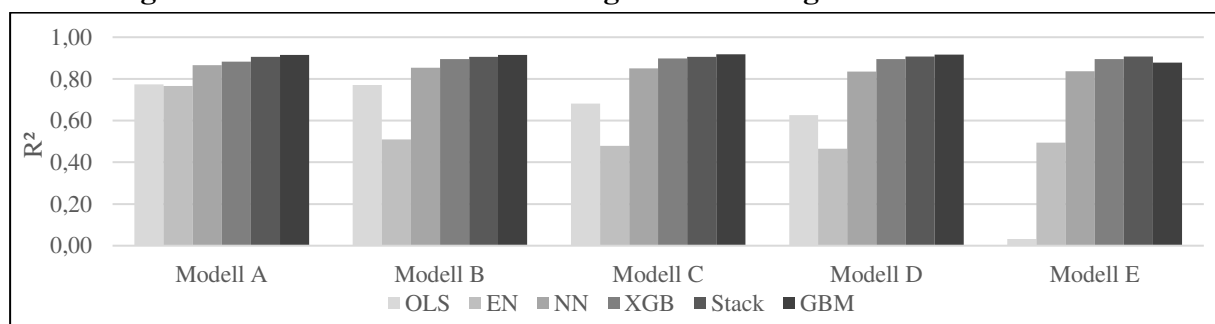
Für die Auswertung der verschiedenen Algorithmen und Modelle werden die trainierten Algorithmen auf den Testdatensatz angewendet. Das bedeutet, dass die monatliche Milchproduktion für alle Betriebe für die jüngsten Beobachtungen von 1/2018 bis 9/2019 vorhergesagt wird. Dies entspricht einem Zeithorizont von 21 Monaten. Durch diesen Ansatz können verschiedene Werte für das relative und absolute Maß der Genauigkeit der Modelle beobachtet werden. Daher wird das Bestimmtheitsmaß³ (R^2), um den relativen Fehler zu messen (siehe Abbildung 3), und der mittlere quadratische Fehler („Root Mean Square Error“ - RMSE) für einen absoluten Fehlerwert, der in Abbildung 4 dargestellt ist, verglichen.

Abbildung 3 zeigt die Unterschiede von R^2 im Vergleich zwischen den Algorithmen, aber auch den Einfluss der Feature Auswahl durch den Unterschied zwischen den Modellen. Die Bezeichnung der Modelle entspricht den in Tabelle 1 angegebenen Abkürzungen. Es fällt auf, dass die beiden linearen Ansätze der OLS und des EN im Verhältnis zu den nicht-linearen Ansätzen ein deutlich niedrigeres Bestimmtheitsmaß (maximal 0,77 für OLS in Modell A) aufweisen. Dieser Unterschied sticht besonders in Modell E hervor, in dem alle Variablen

³ Die Betrachtung des korrigierten R^2 ist nicht notwendig, da in diesem Fall nicht vordergründig die Erklärungskraft durch spezifische Variablen dargestellt werden soll, sondern die Prognosegenauigkeit des gesamten Modells ohne Berücksichtigung von Freiheitsgraden.

berücksichtigt wurden. Das lineare Modell hat mit etwa 3,2 % keine Erklärungskraft für die vorhergesagte Milchproduktion im Test-Set, während das R^2 im Trainingsset bei etwa 0,87 liegt. Dies deutet darauf hin, dass das Modell durch eine zu große Anzahl an Variablen eine Überanpassung vornimmt und dadurch bei neuen Daten aus dem Test-Set keine bzw. nur eine schlechte Vorhersage treffen kann. Dies zu verhindern, ist die Aufgabe des EN, welches durch einen Strafterm weniger relevante Variablen geringer gewichtet. Die Ergebnisse in Abbildung 3 lassen darauf schließen, dass dieser Strafterm bei den Modellen A bis E stark ins Gewicht fällt und es somit zu einer Unteranpassung kommt. Im Vergleich der nicht-linearen Ansätze untereinander zeigen sich nur leichte Unterschiede im R^2 für die jeweiligen Algorithmen und auch zwischen den Modellen A bis E. Mit Ausnahme des Modells E, kann der Algorithmus GBM das höchste Bestimmtheitsmaß (maximal 0,92 in Modell C) vorweisen. Vergleicht man diesen mit dem schematisch ähnlichen Algorithmus XGB, sind nur geringe Unterschiede von bis zu max. 3 % R^2 zu erkennen. Dies erklärt sich durch die gleiche Hyperparameter-Auswahl und den ähnlichen Lernmechanismus. Auf ein Tuning der Hyperparameter wird in diesem Fall bewusst verzichtet, um die Modelle möglichst einheitlich miteinander vergleichen zu können. Das neuronale Netz weist in allen Modellen jeweils das geringste Bestimmtheitsmaß der nicht-linearen Algorithmen auf. Die Kombination der Modelle als Stacking-Algorithmus kann keine nennenswerte Verbesserung im Hinblick auf das R^2 der weiteren Algorithmen vorweisen.

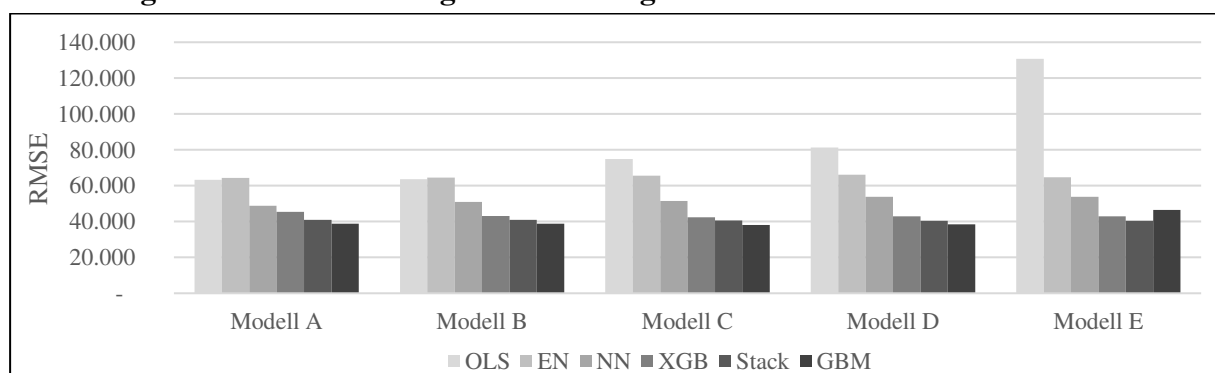
Abbildung 3: Bestimmtheitsmaß für alle angewendeten Algorithmen und Modelle



Quelle: Eigene Darstellung.

Die Ergebnisse des RMSE sind ähnlich zu den zuvor dargestellten R^2 . In Abbildung 4 ist zu erkennen, dass auch im beobachteten RMSE ein Unterschied zwischen linearen und nicht-linearen Ansätzen besteht. Im Gegensatz zum relativen Messwert des Bestimmtheitsmaßes liegen jedoch hier der RMSE der OLS und des EN näher beieinander. Betrachtet man den Wert der linearen Regression für das Modell E, wird auch in Abbildung 4 das Problem der Überanpassung ersichtlich. Abbildung 4 bestätigt, dass es nur geringe Unterschiede zwischen den nicht-linearen Algorithmen gibt. Dabei fällt das RMSE des NN etwas heraus, da hier der Abstand zu den anderen Modellen relativ gesehen größer zu sein scheint als im R^2 -Vergleich und etwa im Mittel zwischen den Werten der linearen und nicht-linearen Modellen liegt.

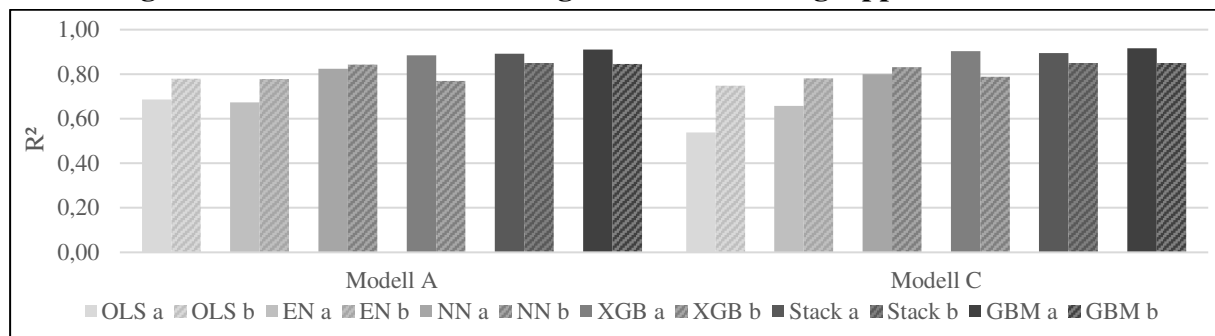
Abbildung 4: RMSE für alle angewendeten Algorithmen und Modelle



Quelle: Eigene Darstellung.

Eine Begründung für den Unterschied der Ergebnisse für R^2 und RMSE in linearen und nicht-linearen Ansätzen könnte die Aufteilung des Datensatzes durch die Indikatorfunktion in moderat und stark wachsende Betriebe sein. Abbildung 5 bildet zur Betrachtung dieses Zusammenhanges das Bestimmtheitsmaß für die jeweiligen Gruppen beispielhaft für die Modelle A und C ab. Die nebeneinander gruppierten Säulen zeigen auf, dass in sämtlichen Kombinationen von Modell und Algorithmus jeweils das R^2 der Gruppe der stark wachsenden Betriebe (Gruppe a) in den linearen Ansätzen (OLS, EN) niedriger ist. Insbesondere im Vergleich zu nicht-linearen Ansätzen fällt das R^2 in diesen jeweiligen Gruppen mit einem maximalen Wert von 0,68 (OLS a, Modell A) deutlich geringer aus. Grund dafür könnten Investitionen als treibender Faktor des Wachstums in Gruppe a sein. Die Milchproduktion in dieser Gruppe steigt nicht linear, sondern teilweise in Sprüngen durch Investitionen in Stallneubauten und/oder -erweiterungen. Diese Entwicklung kann besser durch einen nicht-linearen Algorithmus erklärt werden. So zeigen diese eine gegensätzliche Verteilung, indem die Gruppe der stark wachsenden Betriebe im Vergleich zu der anderen Gruppe besser vorhergesagt werden kann. Dabei liegt das minimale Bestimmtheitsmaß der „Gruppe b“ bei 0,77 (XGB b, Modell A) und damit oberhalb des maximalen Wertes aus der Vergleichsgruppe der linearen Modelle.

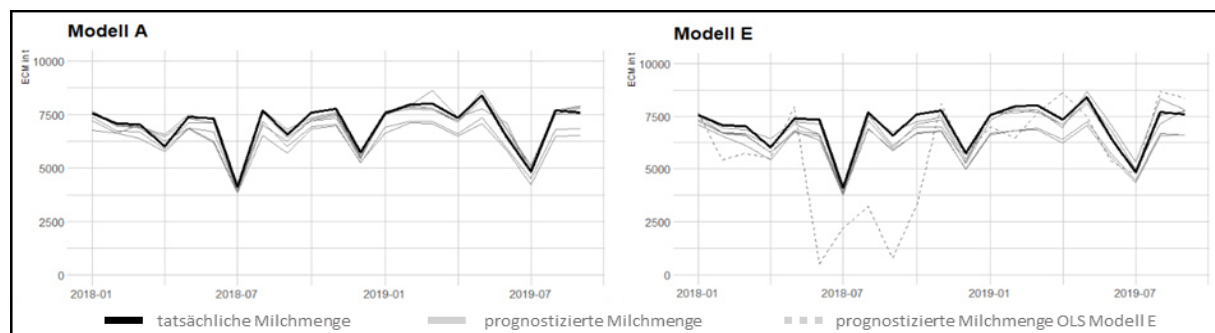
Abbildung 5: Bestimmtheitsmaß im Vergleich der Betriebsgruppen a und b



Quelle: Eigene Darstellung.

Eine Summierung der beobachteten und vorhergesagten Werte der Milchproduktion ermöglicht für den Datensatz eine graphische Betrachtung der tatsächlichen und prognostizierten Milchmenge im zeitlichen Verlauf. In Abbildung 6 ist daher die monatliche Milchmenge in ECM t für den Zeitraum 1/2018 bis 9/2019 dargestellt. Dabei gilt zu beachten, dass für jedes Jahr und Betrieb nur 11 Datenpunkte zur Verfügung stehen. Erhebliche Einbrüche der Milchmenge wie im Monat 7/2018 können auch auf diese Datenverteilung zurückgeführt werden.

Abbildung 6: Vorhergesagte Werte in Summe im Testzeitraum (1/18 – 9/19)



Quelle: Eigene Darstellung.

Die Grafik zeigt, dass die vorhergesagten Werte von Modell A und E sich in ihrer Summe nur sehr geringfügig unterscheiden. Die tatsächlich beobachtete Milchmenge ist als schwarze Linie dargestellt, während die Vorhersage der Modelle in grau abgebildet ist. Auch hier sind die

Unterschiede innerhalb der Modelle zwischen den Algorithmen gering. Auffällig ist hingegen der Verlauf der OLS Regression in Modell E (durchbrochene Linie). Wie schon in den Ergebnissen von R^2 und RMSE zeigt sich hier, dass das Modell durch zu viele Variablen überangepasst ist und dadurch für eine Vorhersage unbrauchbar wird. Interessant ist, dass alle Modelle eher dazu neigen, die Milchmenge zu unterschätzen. Insgesamt kann aber festgestellt werden, dass die angewendeten Modelle jeweils den Trend der Milchproduktion in etwa abbilden.

Ein weiterer Ansatz, um die Modelle zu vergleichen, ist die benötigte Zeit für das Training der Algorithmen. Dies ist insbesondere interessant für eine zukünftig mögliche Erweiterung des Datensatzes um eine größere Anzahl Betriebe. Die Unterschiede zwischen den nicht-linearen Ansätzen werden durch einen Vergleich der Trainingszeit der Modelle in Tabelle 2 ersichtlich. Hierbei liegt der XGB-Algorithmus mit durchschnittlich etwa 47,5 Sekunden Trainingszeit für die jeweiligen Modelle deutlich vor dem NN, dem GBM und der Stacking-Methode, die etwa die 30- bis 55-fache Trainingszeit im Durchschnitt benötigen. Das Stacking dauert auch deshalb länger als die übrigen Methoden, da es die Trainingszeiten der kombinierten Methoden und die eigene Ausführungszeit addiert. Die Trainingszeiten der linearen Modelle sind hingegen mit etwa 4,6 Sekunden (EN) bzw. 0,5 Sekunden (OLS) erheblich schneller.

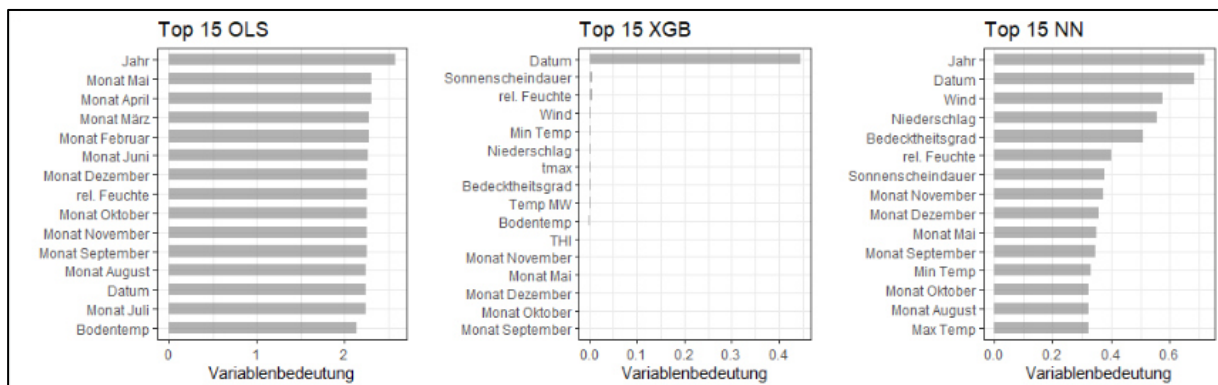
Tabelle 2: Trainingszeit der Algorithmen in Sekunden

	OLS	EN	NN	XGB	Stack	GBM
Modell A	0,05	0,43	387,30	12,70	571,74	184,01
Modell B	0,05	1,63	405,00	16,15	606,74	200,11
Modell C	0,20	1,94	840,42	27,86	1814,90	972,54
Modell D	0,28	2,30	934,50	32,68	2037,62	1100,82
Modell E	0,92	7,52	1263,12	53,07	3072,74	1802,10
Mittelwert	0,50	4,61	1276,78	47,49	2701,25	1419,86

Quelle: Eigene Darstellung.

Zusätzlich zur Vorhersagegenauigkeit und der Geschwindigkeit der Modelle ist ebenfalls von Interesse, welche Variablen relativ den größten Einfluss auf die Vorhersage der Milchmengenproduktion haben. Es bestehen unterschiedliche Verfahren, um die Bedeutung der jeweiligen Variablen zu messen. In diesem Fall wird ein standardisiertes Verfahren der Methode der partiellen Abhängigkeit nach GREENWELL et al. (2018) verwendet. Stellvertretend für die verwendeten Schätz-Methoden werden OLS, XGB und NN verglichen. Bei der Berechnung der Variablenbedeutung für OLS entsprechen die Werte den absoluten Werten der t-Statistik. Für XGB wird die prozentuale Zunahme des MSE errechnet, wenn eine bestimmte Variable in das Gesamtmodell mit einbezogen wird. Die Variablenbedeutung eines neuronalen Netzes wird anhand der Gewichtung der verbundenen Neuronen errechnet. Abbildung 7 stellt einen Vergleich dieser drei oben genannten Algorithmen für Modell A dar. Der Übersichtlichkeit halber werden nur die ersten 15 Variablen nach Variablenbedeutung aufgelistet und die Bedeutung der Betriebsdummies ausgeschlossen. Die absoluten Werte der jeweiligen Algorithmen können nicht miteinander verglichen werden, stattdessen wird die relative Verteilung zueinander betrachtet. Dabei fällt auf, dass für alle Algorithmen zu allererst eine der Zeitvariablen bedeutsam sind. Insbesondere für das XGB-Modell hat die Variable „Datum“ erheblichen Einfluss auf die Vorhersage im Vergleich zu den restlichen Variablen. Diese entspricht dem Zeittrend durch die Kombination aus den Variablen „Jahr“ und „Monat“. Für die OLS Regression sind vor allem die Variablen „Jahr“ und die verschiedenen Monate von höherer Bedeutung. Die Variable „Jahr“ ist ebenfalls für das NN von hoher Relevanz für die Vorhersage, zudem sind einige Wettervariablen wie bspw. „Wind“ und „Niederschlag“ an vorderster Stelle. Damit unterscheidet sich das NN in seiner Reihenfolge der Variablenbedeutung etwas von der OLS und in seinem relativen Verhältnis vom XGB. In allen drei Fällen haben die Milch- und Futterpreisvariablen nur vergleichsweise geringe Bedeutung als Prädiktoren innerhalb der Modelle.

Abbildung 7: Bedeutung der Variablen in Modell A für die Vorhersage



Quelle: Eigene Darstellung.

Bei Betrachtung von Abbildung 7 ist jedoch zu beachten, dass der Vergleich von Feature-Bedeutung nicht gleichzustellen ist mit der statistischen Signifikanz wie in einem linearen Regressionsmodell. Die in Abbildung 7 aufgelisteten Werte sind zwar für die Vorhersage von Bedeutung, sind aber kein präzises Maß für den Einfluss der Variablen auf die Ergebnisse. So kann eine Variable durch eine hohe Korrelation zu einer anderen Variablen (bspw. Minimum und Maximum der Temperatur) eine schlechte Vorhersagebedeutung haben, trotzdem aber eine gewichtete Determinante für die abhängige Variable sein (IFFT et al., 2018).

5 Schlussfolgerungen

Der Vergleich unterschiedlicher Methoden des ML untereinander und im Vergleich zu OLS Regressionen geben Aufschluss, welche dieser Methoden sich für eine Prognose der regionalen Milchproduktion besonders eignen kann. Es werden vier nicht-lineare Ansätze (XGB, GBM, NN, Stacking) und zwei lineare Ansätze (OLS, EN) gegenübergestellt. Der Unterschied zwischen beiden Ansätzen zeigt sich sowohl in relativen als auch in absoluten Indikatoren zur Modellselektion: Die betrachteten nicht-linearen Algorithmen weisen jeweils höhere Bestimmtheitsmaße und geringere RMSE auf. Dabei ist die Spezifikation der Modelle im Hinblick auf Variablen- und Lag-Struktur weniger bedeutend für die Prognosegüte als die Wahl des Schätz-Ansatzes. Die Gegenüberstellung der Variablenbedeutung für die Vorhersage der Modelle verdeutlicht, dass die geringen Unterschiede zwischen den Modellen A bis E aus der höheren Bedeutung von Zeit- bzw. Trendvariablen gegenüber Wettervariablen resultieren. Diese beinhalten einerseits nicht erfasste Wettereffekte und andererseits Management-Entscheidungen, die mit den vorliegenden Daten nicht direkt berücksichtigt werden können. Der Zeittrend kann speziell für die Gruppe der stärker wachsenden Betriebe besser durch nicht-lineare Ansätze erklärt werden. Der Vergleich der nicht-linearen Algorithmen im Hinblick auf die Genauigkeit der Vorhersagen untereinander weist nur im Hinblick auf die benötigte Zeit für das Trainieren der Modelle einen Unterschied auf. Dabei gilt es zwischen einer höheren Genauigkeit (GBM) und einer schnelleren Rechengeschwindigkeit (XGB) abzuwägen. Diese Entscheidung wird besonders bei Prognosen auf Basis großer Datensätzen sehr relevant. Abschließend bleibt festzuhalten, dass ein entsprechend großer Datensatz in Verbindung mit den genannten Algorithmen eine geographische Analyse für das gesamte Bundesland Niedersachsen grundsätzlich ermöglicht, sodass nicht nur die Milchmenge einzelner Betriebe, sondern auch für einzelne Gemeinden und Landkreise prognostiziert werden könnte. Hierzu könnte die verwendete Indikatorfunktion problemlos weitere Effekte des regionalen Strukturwandels berücksichtigen.

Literatur

- ADEDIRAN, S., RATKOWSKY, D., DONAGHY und D., A. MALAU-ADULI (2012): Comparative evaluation of a new lactation curve model for pasture-based Holstein-Friesian dairy cows. In: *Journal of Dairy Science* (95.9), S. 5344-5356.
- ATHEY, S. und G.W. IMBENS (2019): Machine Learning Methods Economists Should Know About. In: *Annual Review of Economics* 2019 (11.1), S. 685-725.
- BLE – Bundesanstalt für Landwirtschaft und Ernährung (2020): Bericht zu Markt- und Versorgungslage mit Milch und Milcherzeugnissen. Berlin.
- BMEL – Bundesministerium für Ernährung und Landwirtschaft (2020): Preise für konventionell erzeugte Kuhmilch. Verschiedene statistische Monatsbericht, 2010-2020.
- BREIMAN, L. (2001): Random Forests. In: *Machine Learning* (45), S. 5–32
- BREIMAN, L., FRIEDMAN, J.H., OLSHEN, R.A. und C.J. STONE (1984): *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.
- FAOSTAT (2020): Statistical database of the Food and Agriculture Organization of the United Nations. Rome, URL: <http://www.fao.org/faostat/en/#home>.
- GOODFELLOW, I., BENGIO, Y. und A. COURVILLE (2016): *Deep Learning*. MIT Press, Cambridge, MA.
- GREENWELL, B., BOEHMKE, B. und A. MCCARTHY (2018): A simple and effective model-based variable importance measure. arXiv preprint:1805.04755.
- HARDER, I., STAMER, E., JUNGE, W. und G. THALLER (2019): Lactation curves and model evaluation for feed intake and energy balance in dairy cows. In: *Journal of Dairy Science* (102.8), S. 7204-7216.
- HASTIE, T., TIBSHIRANI, R. und J. FRIEDMAN (2017): *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, NY.
- HILL, D. und E. WALL (2017): Weather influences feed intake and feed efficiency in a temperate climate. In: *Journal of Dairy Science* (100.3), S. 2240-2257.
- HOEHL, S. und S. HESS (2021): Liquidity, hedging and the survival of North German dairy farms. In: *European Review of Agricultural Economics* (00), S. 1-28.
- IFFT, J., KUHN, R. und K. PATRICK (2018): Can machine learning improve prediction – an application with farm survey data. In: *International Food and Agribusiness Management Review* (21.8), S. 1083-1098.
- JAMES, G., WITTEN, D., HASTIE, T. und R. TIBSHIRANI (2017): *An Introduction to Statistical Learning*. Springer, New York, NY.
- LARCHER, M., SCHÖNHART, M., SCHMID, E. und S. VOGEL (2015): Intensivierungspläne österreichischer MilchproduzentInnen angesichts der auslaufenden Milchquotenregelung 2015. In: *German Journal of Agricultural Economics* (64.03), S. 148-168.
- LSN – Landesamt für Statistik Niedersachsen (2020): Statistische Berichte Niedersachsen, CIII 1 – hj 1 / 2020.
- MADER, T., DAVIS, M. und T. BROWN-BRANDL (2006): Environmental Factors Influencing Heat Stress in Feedlot Cattle. In: *Journal of Animal Science* (84.3), S. 712-719.
- MULLAINATHAN, S. und J. SPIESS (2017): Machine Learning: An Applied Econometric Approach. In: *Journal of Economic Perspectives* (21.2), S. 87-106.
- SCHAPIRE, R. (1990): The Strength of Weak Learnability. In: *Machine Learning* (5), S. 197-227
- SJAUNJA, L.O., BAEVRE, L., JUNKKARINEN, L. und J. PEDERSEN (1990): A Nordic proposal for an energy corrected milk (ECM) formula. 27th Session of ICRPMA, S. 156-157.
- STORM, H., BAYLIS, K. und T. HECKELEI (2020): Machine learning in agricultural and applied economics. In: *European Review of Agricultural Economics* (47.3), S. 1-44.
- VAN DER LAAN, M., POLLEY, E. und A. HUBBARD (2007): Super Learner. In: *Statistical Applications in Genetics and Molecular Biology* (6).
- WOLPERT, D. (1992): Stacked generalization. In: *Neural networks* (5.2), S. 241-259.
- ZOU, H. und T. HASTIE (2005): Regularization and variable selection via the elastic net. In: *Journal of the royal statistical society* (67.2), S. 301-320.