

Big Data Analytics in der Tierwohldebatte

Zwischenstand im Projekt „Landwirtschaft 4.0: Info-System“

Christa Hoffmann¹ und Martin Rieckert²

Abstract: Big Data Analytics ermöglicht Informationen aus Daten automatisch, objektiv und kostengünstig zu extrahieren. So können Daten zur Haltungsumgebung (z. B. Fütterungs- oder Temperaturdaten), aber auch Daten aus Verhaltensbeobachtungen mittels Videokameras oder RFID, analysiert und zur Verbesserung des Tierwohls eingesetzt werden. Eine besondere Bedeutung spielen Maschinelle Lernverfahren, die aus bestehenden Datenbeständen lernen und somit die Datenanalyse vereinfachen, Prognosen für Tierwohl-Risiken ermöglichen und Einflussfaktoren auf das Tierwohl identifizieren. Im Projekt „Landwirtschaft 4.0: Info-System“ werden neue Techniken, Methoden und Verfahren für die intelligente Auswertung entwickelt, um eine breite Zustimmung der Gesellschaft zur wettbewerbsfähigen Tierproduktion zu ermöglichen.

Keywords: Big Data, Maschinelle Lernverfahren, Tierwohl, Schweinehaltung, Data Warehousing

1 Einleitung

Schwanzbeißen ist eines der aktuell viel diskutierten Themen in der Tierwohldebatte. Unbestritten ist, dass die Ursachen für diese Verhaltensstörung multifaktoriell (u.a. Fütterung, Temperatur) sind [EF07]. Es bietet daher eine gute Möglichkeit für den Einsatz von Big Data Analytics. Bereits heute werden eine Vielzahl von Prozessdaten in technischen Anlagen im Stall (z.B. Fütterung) und auch tierspezifische Daten (z.B. Fressaktivität) automatisiert erfasst und als teilweise rein beschreibende, teilweise regulierende Systeme für das Management auf schweinehaltenden Betrieben genutzt [Bml17].

Grundsätzlich machen es die Fortschritte im Bereich des Maschinellen Lernens bereits heute möglich, bisher unbekannte Zusammenhänge in Daten zu identifizieren, zu analysieren sowie Prognosemodelle zu erstellen [Ha09].

Komplexe Zusammenhänge innerhalb und zwischen den einzelnen Datensätzen (z.B. zwischen Trinkverhalten und Abteileremperatur), können derzeit allerdings nur vereinzelt und mit großem Aufwand hergestellt werden. Schnittstellen-Probleme, Medienbrüche und In-sellösungen stellen weiterhin eine Herausforderung dar. Der Großteil der Daten verbleibt ungenutzt als Rohdaten im System.

¹ Bildungs- und Wissenszentrum Boxberg, LSZ, Seehöfer Straße 50, 97944 Boxberg-Windischbuch, christa.hoffmann@lsz.bwl.de

² Universität Hohenheim, Lehrstuhl Wirtschaftsinformatik 2, Schwerzstraße 35, 70599 Stuttgart, martin.rieckert@uni-hohenheim.de

Ziel des Projektes „Landwirtschaft 4.0: Info-System“ ist die Konsolidierung der verteilten und heterogenen Datenbestände in eine zentrale Datenplattform. Die Datenplattform soll Daten über die in den individuellen Einrichtungen (Stallungen, Futterplätze, etc.) verteilten Sensoren, Anlagen, der Managementsoftware (u.a. Sauenplanerdaten) und manuell erfasste Daten aus den Stallungen integrieren, speichern und der Analytik, u.a. im Hinblick auf die Tierwohldebatten, zugänglich machen.

2 Material und Methoden

Durchgeführt wird das Projekt am Standort Bildungs- und Wissenszentrum (LSZ) Boxberg. Die LSZ verfügt über ca. 3.500 Tierplätze (Schweine) in konventionellen und alternativen Bauweisen und einem angeschlossenen Schlachthof. Durch die Auswahl dieser Versuchsanstalt liegen Datenbestände aus dem Routinebetrieb (z.B. Fütterungs-, Temperatur-, Sauenplaner- oder Schlachtdaten) vor. Darüber hinaus kann auf Datenbestände aus speziell verwendeter Versuchstechnik (z.B. Video, RFID, Bewegungslogger) aus aktuellen und abgeschlossenen Versuchen zurückgegriffen werden. Die unterschiedlichen Bauweisen und vielfältige Versuchsansätze zur Tierwohlthematik (u.a. diverse Langschwanzversuche) bieten eine ausreichend komplexe Datengrundlage für die Umsetzung und Potenzialanalyse von Big Data Analytics in der Schweineproduktion.

Um die strukturierten und unstrukturierten Datenbestände für die Datenanalyse zu erschließen, müssen diese Datenbestände zuerst in eine Datenplattform überführt werden (Data Warehousing) [Ba08]. Hierfür wird die Erfassung der am Standort vorhandenen Datenquellen und -bestände, sowie die Erstellung eines Datenbankmodells erforderlich. Dazu wurden am Standort Experteninterviews (Einzel- und Gruppeninterviews) durchgeführt. Die Experten definieren Grenzwerte und Bedingungen, um eine korrekte Datenüberführung zu garantieren (Plausibilisierung).

Daraufhin können Methoden zur Datenanalyse und des Maschinellen Lernens auf dem verknüpften Datenbestand eingesetzt werden [Mi97]. Die Exploration der Daten kann mit Principal Component Analysis [Ab10], Entscheidungsbäumen [Qu86] und unüberwachten, maschinellen Lernverfahren (z.B. Clustering [Ha79]) durchgeführt werden. Des Weiteren ermöglichen überwachte Maschinelle Lernverfahren die Schaffung von Prognosemodellen mit hoher Genauigkeit [Ha09]. Darüber hinaus dienen multivariate Verfahren der Entdeckung und Überprüfung von Zusammenhängen.

3 Ergebnisse

Im bisherigen Projektverlauf ist das Data Warehousing erfolgt. Es kann auf einen verknüpften Datenbestand mit sekundengenauen Messungen für zahlreiche Geräte zurückgriffen werden. Insgesamt sind an der LSZ über 15 verschiedene Geräte und unverknüpfte Datenbestände im Betrieb.

Die Daten lassen sich grob in drei Kategorien einteilen:

1. Biologische Daten (z.B. Lebendgewichte, Schlachtdaten, Sauenplanerdaten)
2. Daten zum Tierverhalten (z.B. Video, RFID, Bonitur)
3. Daten zur Haltungsumgebung (z.B. Temperatur, Lüftung, Energie, Wasser, Futter).

Ein Großteil der Daten kann über CSV-, SQL- oder XML-Schnittstellen aus den Geräten exportiert werden. Ein Teil der Datenbestände, welche über Sensoren erfasst werden (u.a. Wasserdurchflusszähler, Bewegungslogger) kann zentral über eine PRNetBox abgerufen werden. Darüber hinaus sind weitere Datensätze mit Inhalten zur Tiergenetik (Zuchtprogramm) sowie mit ökonomischen Inhalten (u.a. Einkauf-, Verkaufspreise) vorhanden. Lokal abgelegte Exceldateien mit unterschiedlichen Inhalten (u.a. Versuchsauswertungen Fütterung) runden neben vereinzelt handschriftlichen Aufzeichnungen den Datenbestand für die zu erstellende Datenbank ab.

Die Architektur des Infosystems basiert auf einer relationalen Datenbank. Die Daten werden eventbasiert in der Datenbank gespeichert. Vor dem Export findet regelmäßig eine Plausibilisierung der Rohdaten statt. Darauf aufbauend können Daten aus der Datenbank über SQL exportiert und importiert werden und eine Dateneingabe kann über eine Website erfolgen. Die Daten werden durch Reports dem Nutzer angezeigt (Abb. 1).

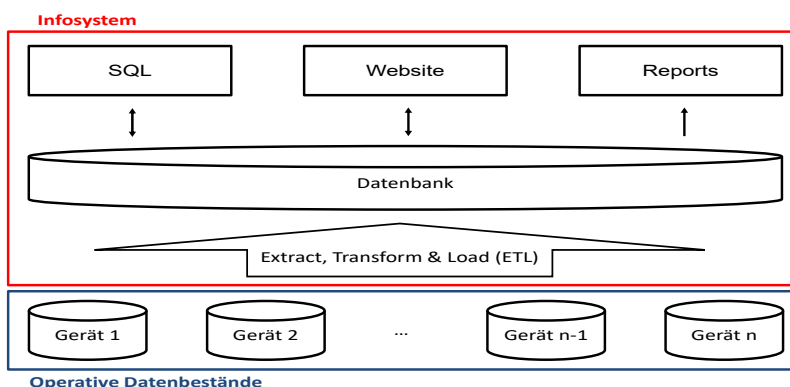


Abb. 3: Architektur des Infosystems

4 Fazit und Ausblick

Im Projekt „Landwirtschaft 4.0: Info-System“ werden neue Techniken, Methoden und Verfahren zur intelligenten Auswertung für die Schweinehaltung entwickelt. Grundlage dafür bietet die bis dato implementierte, solide Datenbankarchitektur. Die Verknüpfung, Extraktion und Aufbereitung der Datenbestände führt zu einer Steigerung der Effizienz der bisherigen Datenerhebung. Im weiteren Projektverlauf werden Methoden zur Datenanalyse und des Maschinellen Lernens auf dem verknüpften Datenbestand eingesetzt. Durch die Anwendung dieser Methoden können Rohdaten zielgerichtet ausgewertet werden. Bisher unerkannte Zusammenhänge zwischen einzelnen Daten werden leichter erkannt. So beispielsweise das Zusammenspiel zwischen Einzelfaktoren (Temperatur, Trinkverhalten und Aktivität) und deren Auswirkungen auf das Tierwohl. Dies ermöglicht neue Lösungsansätze im Rahmen der Tierwohldebatten (u.a. Langschwanz) in der Schweinehaltung. Für Praxisbetriebe ist dadurch eine Verbesserung der Produktionsprozesse und eine Vereinfachung des Managements zu erwarten.

5 Danksagung

Das Projekt „Landwirtschaft 4.0: Info-System“ wird finanziert aus Geldern des Ministeriums für Ländlichen Raum und Verbraucherschutz Baden-Württemberg.

Literaturverzeichnis

- [Ab10] Abdi H, Williams LJ.: Principal component analysis. Wiley Interdiscip Rev Comput Stat 2:433–459, 2010
- [Ba08] Baars H, Kemper H-G.: Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework. Inf Syst Manag 25:132–148 , 2008
- [Bm17] BMEL: Digitalpolitik Landwirtschaft - Zukunftsprogramm: Chancen nutzen - Risiken minimieren, 2017
- [EF07] European Food Safety Authority: Scientific Report on the risks associated with tail biting in pigs and possible means to reduce the need for tail docking considering the different housing and husbandry systems, 2007
- [Ha09] Hastie T, Tibshirani R, Friedman J.: The elements of statistical learning. Springer-Verlag, 2009
- [Ha79] Hartigan JA, Wong MA: A K-Means Clustering Algorithm. J R Stat Soc 28:100, 1979
- [Mi97] Mitchell T.: Machine Learning. McGraw Hill, 1997
- [Qu86] Quinlan JR: Induction of Decision Trees. Mach Learn 1:81–106, 1986