

Food Fraud

Lebensmittelverfälschungen: Unerschlossene Datenbestände gezielt und effizient nutzen



Lebensmittelverfälschungen: Unerschlossene Datenbestände gezielt und effizient nutzen

Hintergrund

Eine grundsätzliche Einordnung des Begriffes „Food Fraud“ in juristische und strukturelle Kontexte sowie weitere in diesem Umfeld notwendige Termini wurden im DLG-Expertenwissen „Food Fraud – Lebensmittelverfälschungen: Möglichkeiten und Chancen zur Risikominimierung in komplex vernetzten Wertschöpfungsketten, 11/2018“¹ vorgestellt. Dort wurde erläutert, wie und ggf. mit welchen Absichten Lebensmittelverfälschungen typischerweise erfolgen. Die nachfolgende Publikation ergänzt diese Erkenntnisse durch einen genaueren Blick auf „Graubereiche“ zwischen gerade noch legalen „Optimierungen“ bei der Herstellung von Lebensmitteln, und Grenzüberschreitungen in Bereiche absichtlicher Täuschung oder gar der Lebensmittelkriminalität.

Möglich wird dieser schärfere Blick u. a. durch automatisierte, konsistent strukturierte Aufbereitungen von Daten und deren laufende, durch Data Science unterstützte Auswertung, wofür einige Beispiele im Text vorgestellt werden. Sie machen deutlich, welche Möglichkeiten in systematischer werkzeuggestützter Datenaufbereitung und -analyse für Prävention und Forensik bei Food Fraud liegen. Gerade in diesen Bereichen zeigt sich das enorme Potenzial der Daten, das aktuell in der Praxis noch nicht optimal ausgeschöpft wird.

Wichtig: Die verwendeten Beispiele für Lebensmittelverfälschungen wurden gewählt, weil damit die jeweiligen Sachverhalte besonders geeignet illustriert werden können. Die Auswahl der Beispiele lässt keine verallgemeinernden Aussagen zu Warengruppen oder Marktteilnehmern zu und soll dieses auch nicht insinuieren.

Werkzeuge zur Abwehr von Food Fraud

Die frühzeitige und zuverlässige Aufdeckung von Lebensmittelverfälschungen² bildet einen wesentlichen Baustein von Strategien zur Erkennung und Abwehr von Lebensmittelkriminalität. Verfälschungen werden dabei in der Regel durch Methoden chemisch-physikalischer Laboranalytik sowie zunehmend auch durch qualifizierte Methoden zur sensorischen Bewertung eines „Degree of Difference“ zu spezifizierten Sollzuständen (DoD-Sensorik) ermittelt (z.B. Difference from Control-Test; In-Out Test). Wie im DLG-Expertenwissen „Food Fraud, 11/2018“ erläutert, können beide Herangehensweisen sowohl gerichtet (Überprüfen einer bekannten Verfälschung) als auch ungerichtet (Überprüfen auf Abweichungen von definierten Referenzen und ggf. nachfolgende Spezifizierung erkannter Abweichungen), sowie separat oder gemeinsam zum Einsatz kommen.

Als Kernbestandteile der Erkennung von Verfälschungen bilden die chemisch-physikalische Laboranalytik und die DoD-Sensorik gemeinsam im Wesentlichen den aktuellen „Werkzeugkasten“ zur Erkennung und Abwehr von Lebensmittelkriminalität – dieser ist bisher für einige wichtige Situationen aber noch nicht ausreichend bestückt.

Grenzen heutiger Werkzeuge zur Abwehr von Food Fraud

Um mögliche Grenzen des heutigen „Werkzeugkastens“ zu erkennen, ist es sinnvoll, das Themenfeld Lebensmittelverfälschung als einen Identifikationspunkt für Lebensmittelkriminalität entsprechend zu strukturieren und die sich ergebenden Strukturelemente auf deren Zugänglichkeit für die verfügbaren Werkzeuge zu analysieren. Im DLG-Expertenwissen „Food Fraud, 11/2018“ wurde als eine der möglichen Ansätze zur Einteilung von Verfälschungen eine Quadrantenstruktur vorgestellt, die auf „Mechanismen“ von Manipulationen basiert und daraus resultierende Risikoklassen ableitet:

- Mechanismen: Produktmanipulation oder Prozessmanipulation
- Risikoklassen: Lebensmittel-Sicherheitsgefährdung gegeben/nicht gegeben

Diese Einteilung führte zu einer Matrix mit vier „Verfälschungs-Quadranten“, in welche verschiedene Szenarien für Verfälschungen überschneidungsfrei einsortiert werden können.

1 <https://www.dlg.org/de/lebensmittel/themen/publikationen/expertenwissen-sensorik/dlg-expertenwissen-112018/>

2 Im weiteren Verlauf wird kurz von „Verfälschungen“ gesprochen

Betrachtet man die vier Quadranten hinsichtlich Sensitivität und Spezifität der heute üblichen Werkzeuge zur Detektion typischer Verfälschungen, wird eines deutlich: Laboranalytik und Prozessaudits leisten vielfach gute Dienste beim Nachweis von Verfälschungen – außer in Fällen, die im Quadranten **„Prozessmanipulation ohne Lebensmittelgefährdung“** einzuordnen sind. Szenarien dieses Quadranten sind u. a. geprägt durch:

- begrenzte „Beweiskraft“ der einzelnen Probe aufgrund fehlender statistischer Zulässigkeit möglicher Rückschlüsse auf die Grundgesamtheit, der die Probe entnommen wurde sowie
- Effekte technisch bedingter Messunsicherheiten, insbesondere in der Nähe von Referenzwerten (z. B. Spezifikations- oder Grenzwerten).

<p>Produktmanipulation mit LM-Sicherheitsgefährdung</p> <ul style="list-style-type: none"> • Melamin in Proteinmischen • Erdnuss in Haselnussmehl • Glykol in Wein • Arylamin-Farbstoffe in Gewürzen 	<p>Prozessmanipulation mit LM-Sicherheitsgefährdung</p> <ul style="list-style-type: none"> • „Rework“ nicht verkehrsfähiger Ware • kein Vorlauf bei Destillation: „höhere Ausbeute“ durch Methanolnutzung • Verwendung industrieller Rohwaren (z. B. Fette, Alkohole, Essigsäure)
<p>Produktmanipulation ohne LM-Sicherheitsgefährdung</p> <ul style="list-style-type: none"> • Haselnussöl in Olivenöl • Sägemehl in Gewürzen • Scholle statt Steinbutt • Konventionell für Bio-Ware • Dokumentenfälschung 	<p>Prozessmanipulation ohne LM-Sicherheitsgefährdung</p> <ul style="list-style-type: none"> • grenzwertige Füllmengensteuerung • Glasuranteil bei TK-Ware • Erhöhung Fremdwasseranteil • Reifezeit bei Hartkäse • Ausreizung von Spezifikationstoleranzen • Spiel mit analytischer Messunsicherheit

Abbildung 1: Quadrant möglicher Verfälschungsmethoden bei Lebensmitteln (LM)

Insgesamt ist es in Verfälschungsszenarien dieses Quadranten im besten Falle „schwer“, oft aber sogar faktisch unmöglich, eindeutige „Schwarz-Weiß“-Entscheidungen zu treffen. Die Grenzen heute gängiger Werkzeuge zeigen sich in der Praxis dann beispielsweise in inhaltlich richtigen, aber nicht klar entscheidungsunterstützenden lebensmittelrechtlichen Beurteilungen wie:

- „Der ermittelte pH-Wert von 4,53 **unterschreitet geringfügig** den festgelegten pH-Wert von 4,6-5,0 gemäß **Produktspezifikation**.“
- „Der präparierte Anteil an Cranberries von 5,0 % **weicht geringfügig** von der Produktspezifikation (**5,5 – 10,6 %**) ab.“
- „Wir weisen darauf hin, dass der ermittelte Wert an Zinn mit 220 mg/kg über dem Höchstgehalt von 200 mg/kg gemäß Anhang VO (EG) Nr. 1881/2006 liegt. Unter Berücksichtigung der erweiterten **Messunsicherheit** ist dieser jedoch **nicht eindeutig** überschritten.“
- „In der vorliegenden Probe wurde >Rückstand< in einer Konzentration von 5,9 µg/kg nachgewiesen, die über dem zulässigen Höchstgehalt von 5,0 µg/kg liegt. [...] Unter Berücksichtigung eines **analytischen Streubereichs** von 37 % ist dieser Höchstgehalt jedoch **nicht eindeutig** überschritten.“

Für Manager sind solche Ergebnisse, etwa im Zusammenhang mit Eigenkontrollen, nicht optimal. Da der „Werkzeugkasten“ objektiv gesehen keine darüber hinausgehenden Informationen ermöglicht, kann der Labor-Dienstleister nicht die eigentlich erwartete belastbare Entscheidungsgrundlage liefern. Zwar wird z. B. eine Nicht-Verkehrsfähigkeit nicht festgestellt und auf Mängel, Abweichungen, Anmerkungen im „Graubereich“ detailliert hingewiesen – diese Informationen sind jedoch nur in den wenigsten Fällen systematisch und effizient zur Verbesserung betroffener Produkte oder als Frühwarnindikatoren zur anschließenden Einleitung von „Abwehrmaßnahmen“ nutzbar. Verfälschungen im Quadranten „Prozessmanipulation ohne Lebensmittelgefährdung“ stellen damit den etablierten methodischen „Werkzeugkasten“ vor kaum zu bewältigende Herausforderungen und bieten potenziellen Lebensmittelfälschern eine (noch) relativ risikoarme Lücke, um daraus wirtschaftliche Vorteile zu erlangen. Im Folgenden wird daher vom „dubiosen Quadranten“ gesprochen.

Food Fraud – oder Grauzone?

Die Ausnutzbarkeit des „dubiosen Quadranten“ für Lebensmittelfälschungen ergibt sich im Wesentlichen aus drei Umständen:

- fehlende statistische Aussagekraft von Einzelproben für induktive Schlüsse,
- Fokus der Prüfpläne auf messtechnisch direkt zugängliche wertgebende und wertmindernde Eigenschaften und
- verfahrensbedingte Messunsicherheiten, insbesondere im Bereich relevanter Sollwerte.

Bestimmte Faktoren erhöhen die Wahrscheinlichkeit, dass konkrete Täuschungen bis hin zur Lebensmittelkriminalität unter Ausnutzung der Situation im „dubiosen Quadranten“ stattfinden. So erscheinen Produkte besonders gefährdet, die eine oder mehrere der folgenden Bedingungen erfüllen:

- haben sehr große Handelsmengen,
- sind homogen,
- kommen homogenisiert in den Verkauf,
- sind aus gut dosierbaren Rezepturanteilen zusammensetzbar.

Derartige Produkte eröffnen diverse Möglichkeiten der Verfälschung, bei denen selbst im Falle eines entsprechenden Stichprobenbefundes ein tatsächlicher Vorsatz heute nur schwer nachweisbar ist, z. B.

- Strecken mit arteigenen Bestandteilen (z. B. das Mitvermahlen von Stängelanteilen),
- Verklappen minderwertiger in verkehrsfähiger Ware unter Einhaltung oder knapper Überschreitung etwaiger Grenzwerte sowie
- Ausreizung zugelassener oder tolerierter „technisch unvermeidbarer“ Streuungen bei Spezifikationswerten³.

Die Existenz des „dubiosen Quadranten“ wird ungewollt durch gesetzliche Regelungen oder etablierte Verkehrsauffassungen für bestimmte Produktgruppen teilweise sogar noch begünstigt: So lässt die Fertigpackungsverordnung (FPackV) nach festgelegten statistischen Mustern eine (zufällig auftretende) Unterfüllung zu⁴, während die „Leitsätze für Fleisch und Fleischerzeugnisse“⁵ für Anteile an wertgebenden Bestandteilen an verschiedenen Stellen unspezifizierte Streubereiche angibt (z. B. Abschnitt 1.72 „Der BEFFE-Gehalt [...] beträgt ca. 20 %.“). Definitionsgemäß gefährden Verfälschungen im „dubiosen Quadranten“ nicht die grundsätzliche Sicherheit von Lebensmitteln. Sofern derartige Fälle überhaupt erkannt wurden, sind sie medial kaum präsent.

Dies darf jedoch nicht als Begründung herangezogen werden, den „Fahndungsdruck“ in diesem Quadranten zu senken. Denn zum einen kann erheblicher monetärer Schaden für betroffene Unternehmen entstehen, der in Form „notwendig“ gewordener Preisanpassungen auch Auswirkungen bis zu Endverbrauchern hätte. Zum anderen zeichnet sich klar ab, dass Supply Chain-orientierte Standards wie IFS und BRC zunehmend einen Kontrolldruck auf die Marktteilnehmer aufbauen und Prüf- bzw. Kontrollsysteme verlangen werden, die Verfälschungen aller Kategorien zuverlässig abfangen können, so dass die Frage „Food Fraud oder Grauzone?“ zukünftig von Unternehmen eindeutig zu beantworten ist.

Wie lassen sich also Verfälschungen rechtzeitig erkennen und abwehren, wenn der aktuelle „Werkzeugkasten“ keine klaren Entscheidungskriterien liefern kann? Indem der methodische „Werkzeugkasten“ um das Potenzial der Daten und ihres gezielten systematischen Managements erweitert wird.

Kombination von Laboranalytik, Prozessaudit und Data Science

Die Abläufe innerhalb der Unternehmen sind bei Feststellung einer Nicht-Verkehrsfähigkeit klar auf Verbrauchersicherheit und Abwehr von Schaden ausgelegt. Sie führen zu robusten Sicherheits- und Korrekturmaßnahmen wie z. B. der Sperrung von Waren vor Verkauf bis hin zu öffentlichen Rückrufen bereits in den Verkehr gebrachter Waren.

Laboruntersuchungen und Beurteilungen der Verkehrsfähigkeit von Lebensmitteln hingegen führen nur selten zur Feststellung einer Nicht-Verkehrsfähigkeit: Für nicht prozessierte Lebensmittel liegt die Quote der Beanstandungen mit „nicht verkehrsfähig“ üblicherweise bei weniger als 1 % aller erstellten Prüfberichte. Bei Mischprodukten und prozessierten Lebensmitteln liegt die Quote abhängig u. a. von der individuellen Warengruppe in der Regel auch nur wenig höher und im einstelligen Prozentbereich.

Im Gegensatz dazu ist der Anteil solcher Prüfberichte regelmäßig deutlich im zweistelligen Prozentbereich, die eine grundsätzliche Verkehrsfähigkeit feststellen, aber die lebensmittelrechtlichen Beurteilungen in Form von z. B. eruierten

3 natürlich bevorzugt zu Gunsten des Lieferanten „streuend“

4 https://www.gesetze-im-internet.de/fertigpackv_1981/FertigPackV_1981.pdf, z. B. 1.1.1

5 <http://www.deutsche-lebensmittelbuch-kommission.de/sites/default/files/downloads/leitsaetzefleisch-2.pdf>

Messergebnissen, des Erfüllungsgrads von Spezifikationswerten oder ermittelte sensorische Abweichungen o.ä. kritisch kommentieren.

Die Inhalte lebensmittelrechtlicher Beurteilungen dieser Art von Prüfberichten werden bislang in kaum einem Unternehmen ganzheitlich und klar strukturiert dokumentiert und ausgewertet. Entsprechend fallen Reaktionen auf vermeintlich oder tatsächlich kritische Inhalte in den Beurteilungstexten der Prüfberichte unterschiedlich aus, da sie vom Verhalten einzelner Mitarbeiter u. a. betriebsspezifisch variierenden Faktoren abhängig sind. Richtige Reaktionen im Sinne eines wirksamen Qualitätsmanagements sind folglich eher bemerkenswerte Einzelleistungen als das Ergebnis planvollen strategischen Handelns.

Viele Unternehmen verfügen über extrem große Daten- und Informationsbestände, für deren Beschaffung, Verwaltung und Analyse sie oft sechs- bis siebenstellige Eurobeträge pro Jahr ausgeben. Trotzdem können sie das Potenzial überwiegend nur punktuell zur nachhaltigen Verbesserung und Prävention nutzen – nämlich dort, wo von (dienstleistenden) Laboren Spezifikations-/Verordnungsverletzungen oder Nicht-Verkehrsfähigkeiten festgestellt und faktenbasiert dokumentiert werden. Informationen, die diese harte Restriktion hinsichtlich ihrer Bereitstellungsform nicht erfüllen, werden dagegen meist lediglich mit ungutem Gefühl abgelegt und verkommen so vom potenziellen Datenschatz zur weitgehend wertlosen Datendeponie.

Ursachen hierfür sind zum einen in der verwendeten (Fach-)Sprache zu suchen, denn eine individuelle Formulierung von Kommentaren im Prüfprotokoll erschwert die maschinelle Erschließbarkeit semantischer Inhalte und der eigentlichen Aussagen. Zum anderen sind vorhandene Hemmschwellen Störfaktoren, die verhindern, dass die noch relativ neuen Techniken des maschinellen Lernens in die Analyse- und Management-Prozesse des täglichen Geschäftsbetriebs integriert werden.

Sofern jedoch in Unternehmen entsprechende Projekte gestartet wurden, zeigt sich meistens, dass üblicherweise genutzte Speicherformate und -medien für Prüfberichte (z. B. PDF, Textverarbeitung, Papier, ...) die maschinelle Weiterverarbeitung der Daten und deren Nutzung für ein gezieltes Qualitäts- und Risikomanagement erschweren. Denn erst wenn die relevanten Texte in einer standardisierten Form (Dateiformat) bereitstehen⁶, kann mit der maschinellen Semantik-Analyse begonnen werden. Daneben bestehen auch inhaltlich-sprachliche Herausforderungen in der Grammatik und bei Rechtschreibfehlern, im Fachvokabular und den dienstleisterspezifischen Bezeichnungen von Laborparametern und Einheiten (z. T. bedingt durch verschiedene Computersysteme), im Satz- und Absatzbereich u. a. Um das gesamte Potenzial eines in einer (großen) Menge von Prüfberichten verborgenen Datenschatzes effizient, konsistent und automatisiert wertschöpfend zu erschließen, sind daher weitere Aufbereitungsschritte sinnvoll oder sogar erforderlich. Bewährt haben sich u. a. folgende Maßnahmen:

- Erzeugung übergeordneter semantischer Zusammenhänge, z. B. „Kuh“ ist ein „Tier“, „Milch“ ist von „Kuh“ (es sei denn, es wird ein anderes „Tier“ explizit im Kontext erwähnt) und somit „tierischen Ursprungs“ usw.⁷
- Herstellung von Zusammenhängen zwischen Textbestandteilen und konkret genannten Einzelmesswerten⁸ eines Prüfberichts.
- Bereitstellung konkreter Messwerte, Warnstufen, Sentimentanalysen etc. über mehrere Produkte und größere Zeiträume hinweg⁹.
- Anbindung von Werkzeugen zur maschinellen Auswertung externer Web-Quellen, etwa Warnportalen, Wetterstationen etc.¹⁰
- Anbindung von Betriebs- und Dienstleistungslaboren, die mit Labor-Informations-Management-Systemen (LIMS) arbeiten, welche über Exportfunktionen für Prüfberichtsdaten in CSV- und/oder XML-Formaten verfügen.¹¹

6 Die Bereitstellung darf nicht auf Projektumfelder beschränkt sein, sondern muss im täglichen Betrieb zuverlässig und robust funktionieren und mit den anfallenden Volumina an Prüfberichten zurechtkommen.

7 Zu externen Warnmeldungen können damit automatisiert potenziell betroffene Waren im eigenen Sortiment ermittelt werden.

8 Wenn im Text z. B. von „erhöhter Keimbelastung“ gesprochen wird, kann über die Verbindung zu konkreten Messwerten automatisch geschlossen werden, welche Keime konkret im Text gemeint sind.

9 Dies ermöglicht neben Zeitreihenanalysen die Verbreiterung von Grundgesamtheiten und damit die Nutzung induktiver statistischer Verfahren bei der frühzeitigen Aufdeckung potenziell lebensmittelverfälschender Prozesse.

10 Erweitert das Spektrum vorlaufender Indikatoren über das eigene Unternehmen hinaus, insbesondere, wenn übergeordnete semantische Verknüpfungen hergestellt wurden (z. B. Regen in der Erntezeit => erhöhte Schimmelgefahr bei Getreide => erhöhte Kontaminationsgefahr durch Mykotoxine in Getreideprodukten).

11 Vereinfacht ggf. die Bereitstellung der Prüfberichtsdaten in einheitlicher, maschinell verarbeitbarer Form erheblich.

Selbst wenn nur Teile heute zur Verfügung stehender Techniken des maschinellen Lernens für automatisierte sprachliche (semantische) Auswertungen von Prüfberichten genutzt werden, ergeben sich vielfältige Möglichkeiten, um das Erkennen von Verfälschungsvorgängen gezielt zu optimieren. Denn die Macht und das Potenzial der Daten zeigen sich auch schon in „kleinen“ Werkzeugen.

Praxisbeispiele und Anwendungen des neuen Werkzeugs „Datenanalyse“

Die Möglichkeiten des maschinellen Lernens zur Analyse von Prüfberichtsdaten sind vielfältig. Nachfolgend soll anhand eines relativ einfachen Beispiels der grundsätzliche Nutzen dargestellt werden.

Dazu wird die Warengruppe der Fleisch- und Wurstwaren und hier speziell homogenisierte bzw. gekutterte Produkte im Zustand roh oder gebrüht betrachtet. Für diese Produktklasse mit verarbeitungsbedingt übereinstimmendem Risikoprofil für bestimmte Eigenschaften wurde eine aussagekräftige Anzahl von Prüfberichten aus verschiedenen Quellen retrospektiv ausgewertet und dahingehend untersucht, welche Messgrößen (Parameter) in Beurteilungstexten in negativen Kontexten erwähnt wurden.

Dazu überführte man die in negativen Kontexten genannten Parameter in eine „Tag Cloud“ oder Wortwolke, wobei die Schriftgröße für einen Begriff gemäß der absoluten Anzahl seiner Nennungen über alle Prüfberichte hinweg gewählt wurde (vgl. Abbildung 2).

Eine dabei viel zu selten genutzte Darstellungsform ist deren Erweiterung zu einer dynamisierten Darstellung, mit einem über den Datenbestand bewegbaren „zeitlichen Fenster“. Damit wären die Wortwolken der verschiedenen Prüfberichte in einer zeitlichen Reihenfolge hintereinander geschaltet, so dass die Datenanalyse je nach Fragestellung auf einen Zeitpunkt oder auf einen Zeitraum fokussiert erfolgen kann (vgl. Abbildung 3). Der visuelle Effekt dynamischer Wortwolken stellt anschaulich die Parameterveränderungen dar und bietet so einen intuitiv verständlichen Baustein bei der Implementierung von Frühwarnsystemen in der praktischen Qualitätskontrolle.

So machte die Anwendung dynamisierter Wortwolken deutlich, dass einige der in der eingangs gezeigten Wortwolke¹² enthaltenen Parameter zu jeder Zeit auftauchen und auffällig¹³ sind, während andere nur zeitlich begrenzt gehäuft auftreten.

Die Detailauswertung zeigte, dass seit Herbst 2019 eine starke Korrelation zwischen der Häufung kritischer Hinweise zu „Listerien“ in Fleisch- und Wurstwaren allgemein und dem „Wilke-Fall“^{14 15 16} bestand. Hervorzuheben ist, dass die für die Auswertung verfügbaren Messergebnisse und Prüfberichte keinen statistisch signifikanten Anstieg der Listerienbelastung der untersuchten Produkte erkennen ließen, es gleichwohl aber vermehrt zu kritischen Beurteilungen genau dieser Parameter „im Graubereich“ (Beurteilungstexte) gekommen war – was bei Betrachtung der Wortwolken-Entwicklung klar erkennbar wurde.



Abbildung 2: „Tag Cloud“ negativer Parameter aus Prüfberichten von Fleisch & Wurstwaren (Anmerkung: Farbe und Position eines Begriffs innerhalb der Wortwolke haben rein ästhetischen Charakter und tragen, anders als die Schriftgröße, keine inhaltliche Bedeutung.)

¹² In einem Print-Medium natürlich leider nur als statische Wortwolke darstellbar.

¹³ also mit großer Schriftgröße dargestellt wurden

¹⁴ <https://umwelt.hessen.de/presse/pressemitteilung/bericht-zum-fall-wilke>

¹⁵ https://umwelt.hessen.de/sites/default/files/media/hmuevl/sachstandsbericht_des_landkreises_waldeck-frankenberg_causa-neu_wilke.pdf

¹⁶ https://umwelt.hessen.de/sites/default/files/media/hmuevl/bericht_rp_kassel.pdf

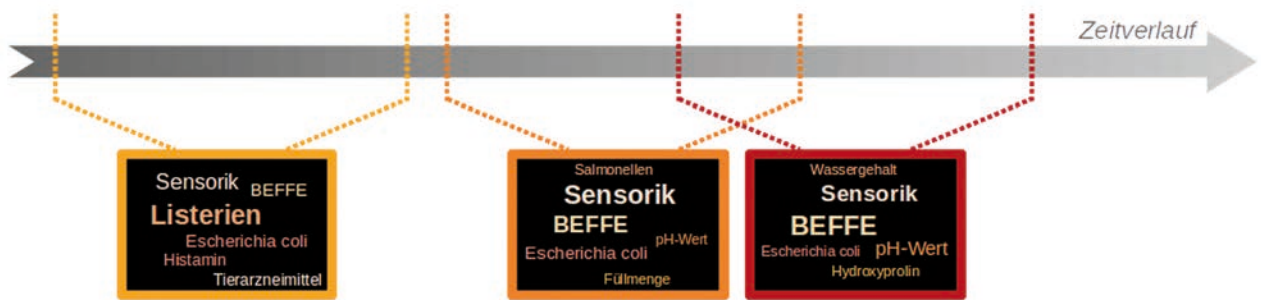


Abbildung 3: „Tag Clouds“ negativer Parameter aus Prüfberichten in dynamischer Darstellung

Diese Methodik ermöglicht z. B. Qualitätsmanagern eines Inverkehrbringers, der diese Klasse von gekutterten Produkten im Sortiment führt und von mehreren Lieferanten bezieht, die in Wortwolken auftauchenden Begriffe als Hinweise auf erhöhte Prüfbedarfe für die eigenen gehandelten Produkte bzw. Lieferanten zu nutzen. Mit herkömmlichen Datenbereitstellungsformen wäre dieser Sachverhalt nicht so einfach zu erkennen. Wird das Potenzial der Daten dagegen geschickt genutzt, finden sich derartige Hinweise innerhalb weniger Mausklicks auch für Nutzer*innen ohne detaillierte Warenkenntnisse.

Kommerziell verfügbare Verfahren zur Datenerschließung, nutzen Werkzeuge der Data Science und Methoden des maschinellen Lernens (kurz: ML), um nach einer „Trainingsphase“ des Systems unternehmensspezifische Aufbereitungen, Harmonisierungen und Anreicherungen¹⁷ der Gesamtheit aller verfügbaren Prüfberichte durchzuführen und die Daten nach der Auswertung entsprechend aufbereitet zur Verfügung zu stellen.

Die folgenden Abbildungen wurden Bildschirmdarstellungen eines Web-Interfaces¹⁸ entnommen und zeigen erste einfache Nutzungsmöglichkeiten. Zur einfachen Unterscheidung von im Prüfbericht festgestellten Risiken wurde anwenderseitig hier eine vertraute „Ampelsystematik“ gewählt.

Entsprechend der Ampelsystematik bedeuten die Farben:

- /grün = keine Beanstandung
- /gelb = kritische Hinweise
- /rot = schwere Beanstandung bzw. Nicht-Verkehrsfähigkeit

Die Vergabe von Ampelfarben kann sich summarisch auf komplette Prüfumfänge stützen, alternativ können die verwendeten ML-Algorithmen aber auch bestimmte Parameter oder -kombinationen isoliert bewerten.

Durch die Anwendung auf alle risikoverwandten Produkte¹⁹ eines Lieferanten lassen sich dann auf der Nutzeroberfläche automatisiert Ampeldarstellungen erzeugen, die systematische Qualitätsthemen leicht erkennen lassen. Für die gezeigte Situation wurden ML-Algorithmen des Systems auf den (in der oben gezeigten zugehörigen Wortwolke am häufigsten vorkommenden) Parameter „BEFFE“ abgestellt. „BEFFE“ steht für „bindegewebs-eiweißreiches Fleischeiweiß“ und repräsentiert den Eiweißanteil des reinen Muskelfleisches als ein Maß für die Qualität einer Fleischware²⁰. Der Anteil an BEFFE in stark homogenisierten Proben wird in der Regel durch Leitsätze geregelt (s. o.) oder näher spezifiziert und lässt sich in der Produktion gut kontrollieren. Die Überprüfung auf den tatsächlichen BEFFE-Gehalt erfolgt im Labor über indirekte Methoden.

Abweichungen in Richtung kleinerer BEFFE-Werte spezifiziert oder wie im Leitsatz gefordert, führen für Einzelproben in der Regel wegen fehlender statistischer Relevanz der Probe zwar zu kritischen Hinweisen ●/gelb aber nur in seltenen Fällen zu einer Feststellung schwerer Abweichungen ●/rot – mithin eine mögliche Situation aus dem grauen Bereich des „dubiosen Quadranten“.

¹⁷ z. B. mit Umfelddaten aus Warnportalen, Wetterberichten, ...

¹⁸ Verwendet wurde hier das Basissystem der Risksniffer GmbH, da die Autoren als Gründer des Unternehmens mit diesem System besonders vertraut sind. Andere Systeme mögen durchaus ähnliche Datenaufbereitungen und Auswertungen erlauben.

¹⁹ Hier kommen u. a. semantische Meta-Informationen zum Einsatz.

²⁰ z. B. <https://www.lebensmittelwissen.de/lexikon/b/BEFFE.php>

Artikel-Nr.	Bezeichnung	Validierung letzter Prüfbericht	Warnstufe letzte Prüfberichte	
			✓ ✓ ✓ ✓ 1 1 1 △ △ △ △ △	▶
			✓ ✓ ✓ ✓ 1 △ ✓ 1	▶
			✓ △ ✓ 1 1 1 △ △ △ △ ✓	▶
			✓ △ △ △ △ △ △ △ ✓	▶
			✓ 1 △ ✓ ✓ ✓ ✓ ✓	▶
			✓ ✓ 1 △ △ △ △ ✓ ✓	▶
			✓ △ ✓ △ △ ✓ ✓ ✓ ✓	▶
			✓ △ 1 △ △ △ ✓	▶
			✓ △ △ ✓ △ △ ✓ ✓ 1 ✓	▶

Abbildung 4: Fallbeispiel 1: Auswertung von Prüfberichten und Warnstufen
(Hinweis: Unleserliche Stellen in den Bildschirmausschnitten resultieren aus Datenschutzgründen)

Ergänzend zur Ampelsystematik wird bei dem gezeigten System auf Schriftzeichen und geometrische Formen (Haken in „grünem“ Kreis, Ausrufezeichen in „gelbem“ Kreis und Ausrufezeichen im „roten“ Dreieck) zurückgegriffen. Dies dient dazu, auch bei Schwarz-Weiß-Ausdrucken der Bildschirminhalte oder für Nutzer*innen mit Rot-Grün-Sehschwäche eine zuverlässige Identifikation zu ermöglichen.

Fallbeispiel 1: Wirkung qualitätsverbessernder Maßnahmen

Mit dieser Systematik werden alle dokumentierten Proben (Laboruntersuchungen mit Prüfberichten) sämtlicher Artikel vergleichbaren Risikoprofils²¹ im definierten Analysezeitraum hinsichtlich BEFFE-Aussagen bewertet. Dazu werden die Proben basierend auf Messwerten, Hinweisen in den natürlich-sprachlichen Beurteilungstexten und systemseitig erzeugten Metadaten ausgewertet und die Ergebniseinschätzung mit dem entsprechenden Symbol zur Anzeige gebracht (vgl. Abbildung 4). Jede Zeile in den Bildschirmausschnitten entspricht einem spezifischen Produkt, für das Beispiel sind dabei nur die Inhalte der Spalte „Warnstufe letzte Prüfberichte“ von Interesse: Die Ergebnissymbole sind chronologisch absteigend sortiert, d.h. ganz rechts steht die älteste Probe/der älteste Prüfbericht aus dem untersuchten Zeitraum, ganz links steht entsprechend die neueste Probe/der neueste Prüfbericht²².

Bereits diese einfache zeitliche Einsortierung der maschinell erzeugten Bewertungen (verdichtet auf die beschriebenen Symbole) zeigt unmittelbar, dass für einige Produkte dieses Risikoprofils kritische Beurteilungen von Beginn an eher die Regel als die Ausnahme waren. Zusätzlich wird anhand der Häufung kritischer Hinweise und schwerer Beanstandungen deutlich, dass die Qualität bei mehreren Produkten dieser Gruppe zwischenzeitlich nicht ausreichend war. Die jeweils aktuellen Prüfberichte (ganz links) sind dann wiederum durchgängig ohne Beanstandung, was als Indikator für die Wirksamkeit getroffener Verbesserungsmaßnahmen interpretiert werden kann. Zur Absicherung dieser Interpretation gilt es, das Monitoring zukünftig in vergleichbarer Intensität aufrechtzuerhalten. Da dies entsprechende Systeme vollmaschinell leisten, lässt sich diese Aufgabe gut in den Arbeitsalltag des Qualitätsmanagements integrieren.

²¹ Die Auswahl wird hier vom System anhand von Metadaten automatisch getroffen, maschinell vorgeschlagene Auswahlen können aber manuell angepasst oder vollständig durch manuell getroffene Auswahlen ersetzt werden.

²² In dieser Darstellung werden für die Gesamtbeurteilung auch hart festgestellte Nicht-Verkehrsfähigkeiten mit als roter Kreis mit Ausrufezeichen angezeigt.

Artikel-Nr.	Bezeichnung	Validierung letzter Prüfbericht	Warnstufe letzte Prüfberichte	
			▲ ▲ ▲ ▲ ▲ ● ● ● ● ● ● ● ● ● ●	▶
			▲ ▲ ▲ ▲ ● ● ● ● ● ● ● ● ● ●	▶
			▲ ▲ ▲ ▲ ▲ ● ● ● ● ● ● ● ● ● ●	▶
			▲ ▲ ▲ ▲ ▲ ▲ ● ● ● ● ● ● ● ● ● ●	▶
			▲ ▲ ▲ ▲ ● ● ● ● ● ● ● ● ● ●	▶
			▲ ▲ ▲ ▲ ▲ ● ● ● ● ● ● ● ● ● ●	▶

Abbildung 5: Fallbeispiel 2: Auswertung von Prüfberichten und Warnstufen
(Hinweis: Unleserliche Stellen in den Bildschirmausschnitten resultieren aus Datenschutzgründen)

Fallbeispiel 2: Aufdeckung von beginnender Qualitätserosion

Fälle schleichender Qualitätserosion können durchaus längere Zeit unentdeckt im „dubiosen Quadranten“ Spuren hinterlassen, die mit modernen Werkzeugen früher hätten identifiziert werden können.

Bei den in Abbildung 5 zusammengestellten und betrachteten Artikeln ist ein für diese Situation typisches Muster erkennbar – sofern verwendete Werkzeuge in der Lage sind, „zwischen den Zeilen“ lebensmittelrechtlicher Beurteilungen zu „lesen“. Nach tadelloser Erfüllung der Erwartungen (Spezifikation, Leitsätze o.ä.) zu Beginn der Belieferung (älteste Prüfberichte, damit rechts in der jeweiligen Zeile ausgewiesen) häufen sich nach einem längeren Zeitraum Hinweise auf Qualitätsmängel, die zwar nicht den Status eines „nicht verkehrsfähig“ erlangen, Labore aber trotzdem dazu bewegen haben, im Berichtstext kritische Hinweise²³ im Zusammenhang mit BEFFE zu geben.

Im Arbeitsalltag der Unternehmen bleibt mit heutigen Werkzeugen eine derartige Veränderung über eine ganze Produktklasse hinweg oft unentdeckt. Dies ermöglicht es Marktteilnehmern, Manipulationen an Lebensmitteln lange sanktionsfrei zu Lasten des Abnehmers und letztlich der Verbraucher vorzunehmen.

Wie im Fallbeispiel 1 zeigt sich auch hier, dass bereits relativ simple ML-basierende Auswertungen deutliche Vorteile im proaktiven Qualitäts- und Risikomanagement bieten können. Trotzdem müssen auch moderne Werkzeuge mit Sachkenntnis verwendet werden.

Effizienzgewinn

Der operative Wert Data Science/ML (Machine Learning)-basierter Auswertungen liegt in Effizienzgewinnen, die durch eine Fokussierung knapper Unternehmensressourcen auf die Überprüfung konkreter Verdachtsfälle erzielt werden anstelle eines großflächigen Stichprobenverfahrens, das – wie bereits erläutert – im „dubiosen Quadranten“ sowieso an enge systemimmanente Grenzen stößt. Diese engen systemimmanenten Grenzen heutiger Werkzeuge werden durch Data Science/ML-basierte Ansätze v. a. dadurch deutlich weiter gezogen, dass trotz eingeschränkter statistischer Signifikanz der Einzelergebnisse, übergeordnet Trends erkennbar werden können, sofern mehrere Artikel eines vergleichbaren Risikoprofils miteinander verglichen werden, wie im Beispiel für den qualitätsanzeigenden Parameter BEFFE.

Wenngleich es in kaum einem Prüfbericht zu einem „harten Urteil“ bezüglich BEFFE kommt, liefert das zeitgleiche wiederholte Auftreten kritischer Anmerkungen diesbezüglich genügend „statistische Evidenz“ für einen Anfangsverdacht, dass dieser Lieferant mit Intention vorgeht oder einen unerkannten Mangel durchreicht.

²³ Sentimentanalysen, wie im hier verwendeten System implementiert, erlauben, automatisch mit hoher Zuverlässigkeit zwischen positiver und negativer Erwähnung von Parametern in Berichtstexten zu unterscheiden.

gehandelten Artikel handelte, der seitens des Zulieferers qualitativ nicht explizit überprüft wurde. Mithin wurden die Mängel vom Zulieferer unerkannt durchgereicht – was beinahe zur Auslistung dieses ansonsten sehr zuverlässigen Lieferanten geführt hätte, wäre der Sachverhalt nicht wie beschrieben aufgeklärt worden.

Für drei weitere Artikel der Liste finden sich im jeweils aktuellsten Prüfbericht simultan kritische Anmerkungen zu BEFFE, was als Hinweis auf eine beginnende Qualitätserosion gewertet wurde. Wie bereits vorher beschrieben, bedarf es in solchen Anfangsverdachtsfällen weiterer Analysen, bevor endgültige Schlüsse zulässig sind. Im vorliegenden Fall wurden im Rahmen von Audits drei wechselseitig unabhängige Ursachen für die kritischen Anmerkungen gefunden, was letztlich dazu führte, dass der Anfangsverdacht auf intentionale Verfälschung fallen gelassen werden konnte.

Zusammenfassung mit Ausblick auf die Wertschöpfungskette

Bereits die wenigen, thematisch auf eine Visualisierungsform und einen Parameter begrenzten Beispiele zeigen, dass die Anwendung von Verfahren der Data Science und des Machine Learning auf Prüfberichtsdaten eines Unternehmens deutliches Potenzial bieten: sowohl bei der präzisen und frühzeitigen Identifikation von Verdachtsfällen, als auch in der Zuweisung von Unternehmensressourcen im Qualitätsmanagement. Gerade im Vergleich mit heute üblichen Auswertungen liegen in der maschinell unterstützten systematischen Umfeldausleuchtung des „dubiosen Quadranten“ große Chancen, Fällen von Food Fraud frühzeitig ggf. bereits in der Vorbereitungsphase zu begegnen.

Die Macht und das Potenzial der Daten geht sehr weit über die in diesem DLG-Expertenwissen gezeigten Möglichkeiten eines Einzelunternehmens hinaus. Die Lebensmittelindustrie verfügt über einen enormen Datenschatz, der mehr oder weniger die gesamte Wertschöpfungskette „Farm-to-Fork“ umfasst. Würde die Branche sich entschließen, diesen Datenschatz gemeinsam geeignet zu bewirtschaften und den Data Science-/ML-Werkzeugkasten gemeinschaftlich zu erweitern, könnten Frühwarnindikatoren bereits in früheren Prozessstufen implementiert und über Verfahren des maschinellen Lernens automatisiert werden. Damit könnten sich verändernde Risikolagen zuverlässig maschinell erkannt und nachgelagerte Prozessteilnehmer frühzeitig gewarnt werden. Zum Beispiel könnten kontaminierte Rohwaren bereits identifiziert werden, bevor (potenziell mehrere) Hersteller diese in eigenen Produkten verarbeitet haben, was die branchenweite Risikoexposition deutlich mindert. Auch das Risiko der Unternehmen, Mängel unerkannt durchzureichen, könnte mit einem übergreifenden Datenmanagement- und Datenanalysesystem stark reduziert werden.

State-of-the-Art Data Science-/ML-Systeme ergänzen unternehmenseigene Daten zwar manchmal bereits um individuell zusammengestellte Umfelddaten, z. B. aus offiziellen Meldesystemen und Frühwarndiensten, Wetterdiensten oder ähnlichem und liefern so wertvolle Beiträge zur Optimierung des Qualitätsmanagements – verglichen mit einem unternehmensübergreifenden Datenpool, der sich aus möglichst vielen Teilnehmern verschiedener Wertschöpfungsstufen speist, werden erzielbare Effekte im Vergleich aber beschränkt bleiben.

So wären wirtschaftliche Vorteile, die sich etwa für Inverkehrbringer aus der Bewirtschaftung eines solchen unternehmensübergreifenden Datenpools ergeben könnten, durchaus substanziell: Produkte mit vergleichbaren Risikoprofilen müssten nicht von jedem Unternehmen individuell auf die gleichen Standardparameter getestet werden, sondern Prüfpläne könnten auf Basis automatisch generierter (anonymisierter) Hinweise aus dem gemeinsamen Datenpool dynamisch im Hinblick auf aktuelle Verdachtssituationen geschärft werden – wenn im Netzwerk „an der BEFFE-Front Ruhe herrscht“. Während über einen (anonymisierten) Teilnehmer des Netzwerks bekannt wird, dass bei ihm in jüngerer Zeit gehäuft Arzneimittelrückstände in Prüfberichten negativ erwähnt werden, wäre eine temporäre Verschiebung eigener Geldmittel von BEFFE- auf Arzneimittelrückstandsanalysen zu überlegen – Qualitätsmanager könnten also ihren Mitteleinsatz mittels evidenzbasierter dynamischer Prüfplangestaltung effizienter gestalten.

Leider finden die genannten Vorteile, die ein sich von möglichst vielen Marktteilnehmern bewirtschafteter Datenpool bietet, noch wenig Resonanz in der Praxis. Dem Konzept einer vernetzten gemeinsamen Nutzung von Datenpools mit Labordaten, Beurteilungen und flankierenden Informationen wird aktuell noch reserviert begegnet. Ursachen hierfür werden hauptsächlich gesehen in:

- Interessenskonflikten verschiedener Marktteilnehmer auf unterschiedlichen Stufen der Wertschöpfungskette sowie
- der verbreiteten Sorge, sich mit der Bereitstellung eigener, teils hochsensibler Daten trotz wirksamer Anonymisierung vor Bereitstellung in einem Datenpool zu exponieren.

Die Umsetzung des Konzeptes eines gemeinsam bewirtschafteten Datenpools der Lebensmittelindustrie bedingt eine enge Zusammenarbeit zwischen den Disziplinen

- Warenkunde (Leitsätze, Spezifikationen, ...),
- Prozesskunde (Laborverfahren und deren Grenzen, Herstellungsprozesse, ...),
- Datengewinnung, -aufbereitung und -speicherung (inhaltlich und technisch, IT-Infrastruktur, ...),
- Data Science (Feature Engineering, Model-Selection und -Training) und
- Recht (Lebensmittelrecht, Datenschutzrecht, IT-Recht, ...).

Insgesamt wird deutlich, dass die Schaffung eines von möglichst vielen Marktteilnehmern getragenen, gemeinschaftlich genutzten Datenpools zur automatisierten Überwachung neuralgischer Punkte der Wertschöpfungsketten bis dato noch visionär ist: Technische und organisatorische Infrastrukturen sind aktuell noch nicht auf eine solche Zusammenarbeit ausgerichtet und Interessenslagen potenzieller Teilnehmer sind in vielen Fällen nicht ausreichend deckungsgleich.

Aus der Nutzung erster kommerziell erhältlicher Produkte und deren Anwendung auf Eigenkontrolldaten und Untersuchungen direkter Vorlieferanten ergeben sich aber bereits bei nur unternehmensinternem Einsatz sichtbare Transparenzgewinne sowie konkrete Ansatzpunkte für systematische Qualitätsverbesserungen, teilweise gemeinsam mit direkten Vorlieferanten.

Schon in den unternehmenseigenen Daten steckt also im Zusammenspiel mit Data Science-/ML-Verfahren Macht:

- Entlastung des operativen Qualitäts- und Risikomanagements von Routinearbeiten
- gezieltere Steuerung personeller und finanzieller Ressourcen der Qualitätssicherung,
- frühere und spezifischere Möglichkeiten zur risikolagenadäquaten Dynamisierung von Prüfplänen,
- flexible Auswertungsmöglichkeiten bei sich änderndem Informationsbedarf.

Vielleicht werden im Zuge eines immer selbstverständlicheren Umgangs der Unternehmen mit solchen Systemen und den erlebten positiven Auswirkungen auch die Chancen steigen, einen branchenweiten gemeinsam bewirtschafteten Datenpool Wirklichkeit werden zu lassen. Erst dann könnte die Branche das Potenzial der Daten vollständig nutzen.

Autoren:

Anke van Hasenhorst, Dr. Andreas Müller, Bernd Lehmkuhl

Risksniffer GmbH, Bargteheide, www.risksniffer.de, Kontakt: Andreas.Mueller@risksniffer.de

Kontakt:

Bianca Schneider-Häder, DLG-Fachzentrum Lebensmittel, Sensorik@DLG.org

© 2020

Alle Informationen und Hinweise ohne jede Gewähr und Haftung. Vervielfältigung und Übertragung einzelner Textabschnitte, Zeichnungen oder Bilder – auch für den Zweck der Unterrichtsgestaltung – nur nach vorheriger Genehmigung durch DLG e.V., Marketing, Eschborner Landstraße 122, 60489 Frankfurt am Main.



DLG e.V.

Fachzentrum Lebensmittel

Eschborner Landstraße 122 · 60489 Frankfurt am Main

Tel. +49 69 24788-311 · Fax +49 69 24788-8311

FachzentrumLM@DLG.org · www.DLG.org