



## BERICHT

zu Forschungsvorhaben

B83 + B83 (Forts.)

Projekt-Nr.

- Zwischenbericht Nr.: \_\_\_\_\_
- Abschlußbericht

für den Zeitraum vom: 01.01.2004 bis: 31.12.2005

### **Themen:**

Optimierung des Filtrationsprozesses in der Brauerei durch Anwendung von Data Mining  
Optimierung des Filtrationsprozesses in ausgewählten Brauereien: Vertiefende Ursachenforschung und Praxisimplementierung

### **Forschungsstellen:**

FS 1: InformationsTechnologie Weihenstephan, Am Forum 1,  
85350 Freising

FS 2: Lehrstuhl für Fluidmechanik und Prozeßautomation, Weihenstephaner Steig  
23, 85350 Freising

### **Leiter der Forschungsstellen:**

FS 1 und 2: Univ.-Prof. Dr.-Ing. A. Delgado

### **Projektleiter:**

Dr.-Ing. Rainer Benning

### **Verfasser des Berichtes:**

Dr.-Ing. Rainer Benning

### **Datum:**

30.03.2006

Bitte Inhaltsübersicht beifügen

<b>1</b>	<b>ZUSAMMENFASSUNG</b> .....	<b>3</b>
<b>2</b>	<b>WISSENSCHAFTLICHE PROBLEMSTELLUNG</b> .....	<b>3</b>
2.1	AUSGANGSSITUATION.....	3
2.2	MOTIVATION FÜR DEN FORSCHUNGSGEGENSTAND.....	4
<b>3</b>	<b>FORSCHUNGSZIEL / LÖSUNGSWEG</b> .....	<b>4</b>
3.1	FORSCHUNGSZIEL UND ANGESTREBTE FORSCHUNGSERGEBNISSE.....	4
3.2	ERGEBNISSE.....	5
3.2.1	<i>Zusammenfassung</i> .....	5
3.2.2	<i>Datenvorverarbeitung</i> .....	6
3.2.3	<i>Auswahl der Data – Mining Verfahren</i> .....	6
3.2.4	<i>Klassifizierung der Filtration</i> .....	7
3.2.5	<i>Filtrationsklassen als Grundlage der Prädiktion</i> .....	8
3.2.6	<i>Modellerstellung</i> .....	9
3.2.6.1	Allgemeines.....	9
3.2.6.2	Künstliche Neuronale Netze.....	9
3.2.6.3	Predict – Tree – Verfahren.....	10
3.2.6.4	Fuzzy Logik.....	10
3.2.6.5	Präzisierung der Bewertung.....	11
3.2.6.6	Untersuchungen zum Einfluss der Reihenfolge der Sorten am Filter.....	12
3.2.6.7	Analyse der Labordaten.....	12
3.2.7	<i>Fazit</i> .....	13
<b>4</b>	<b>AUSBLICK AUF DAS FOLGEVORHABEN</b> .....	<b>14</b>
<b>5</b>	<b>LITERATURLISTE</b> .....	<b>14</b>

# 1 ZUSAMMENFASSUNG

Primäre Ziele der beiden Vorhaben war es, die prinzipielle Anwendbarkeit von Data Mining – Algorithmen auf die Untersuchung der Filtration aufzuzeigen und aus einer darauf aufbauenden Modellierung Ansätze zur Optimierung der Filtration zu entwickeln. Der grundsätzliche Nachweis der Anwendbarkeit von Data Mining Methoden konnte erfolgreich erbracht werden. Mit den daraus gewonnenen Erkenntnissen ließen sich Ansätze zur Vorhersage der Filtrierbarkeit entwickeln. Eine wesentliche Grundlage lag in der ausschließlichen Verwendung von routinemäßig erhobenen Betriebsdaten sowohl als Grundlage zur Bewertung der Filtrationen als auch zur darauf aufbauenden Prädiktion der Filtrierbarkeit. Damit entsteht beteiligten Brauereien zunächst kein zusätzlicher Mess- und Analysenaufwand.

Die Fortsetzung der Forschungsarbeiten zeigte, dass die im ersten Teil entwickelten Ansätze sich noch nicht zur Entwicklung eines Modells ausreichender Prädiktionsgüte und erwünschter Erweiterbarkeit eigneten. Tiefergehende Untersuchungen der vorliegenden Daten führten zu einer modifizierten Modellierungsbasis, die eine getrennte Betrachtung wirtschaftlicher und qualitativer Größen erlaubt.

Als Konsequenz der erzielten Ergebnisse ergab sich die Beantragung eines Folgevorhabens zur Entwicklung eines „virtuellen Filtrationsassistenten“, dessen Aufgabe in der Unterstützung des Filtermeisters durch Aufzeigen von Regelungshinweise besteht, und dessen Beginn zum 1.07.2006 geplant ist.

## 2 WISSENSCHAFTLICHE PROBLEMSTELLUNG

### 2.1 Ausgangssituation

Die Ausgangssituation für die durchgeführten Projekte begründete sich in der Bedeutung der Filtration für die Bierqualität [1, 2] und der Tatsache, daß trotz einer Vielzahl von Untersuchungen auf diesem Gebiet wesentliche Bestimmungsfaktoren noch nicht ausreichend erforscht sind. Insbesondere konzentriert sich die Mehrheit der Forschungsansätze vor allem auf die Inhaltsstoffe des Bieres und ihre Auswirkungen auf die Standzeit der Filter, sowie auf die Filtrationsart und Filterhilfsmittel [z.B. 2, 4, 5].

Obwohl die Anschwemmfiltration mittels Kieselgur derzeit kritisch diskutiert wird [3], stand sie aufgrund ihrer derzeit immer noch zentralen Bedeutung im Mittelpunkt des Projekts. Durch die Zudosierung des Filterhilfsmittels zum Bier und die laufende Aufschwemmung auf eine geeignete Trägerschicht entsteht eine sich ständig erneuernde Oberfläche, die eine kontinuierliche Filtration ermöglicht. Die mit der Zeit stetig anwachsende Dicke der Filterschicht erfordert zur Aufrechterhaltung des Prozesses steigende Druckdifferenzen zwischen Filtereinlauf und -auslauf [4], die jedoch anlagenspezifische Obergrenzen aufweisen.

Ein kostengünstiger Prozess lässt sich unter Sicherstellung der Bierqualität durch eine lange ungestörte Filtration und eine möglichst geringe Kieselgurmenge, inklusive der Reduktion von Hilfsstoff- und Entsorgungskosten, erreichen.

## **2.2 Motivation für den Forschungsgegenstand**

Die grundlegende Motivation der durchgeführten Vorhaben bestand aufgrund der geschilderten Ausgangssituation in der Erstellung eines praxistauglichen Systems zur Prädiktion und Optimierung des Filtrationsprozesses in der Brauerei auf der Basis der Auswertung vorhandener Daten mittels Data Mining.

Aufgrund der jeweils auf ein Jahr beschränkten Projektlaufzeit sollte dabei in einem ersten Schritt die prinzipielle Anwendbarkeit des Data Mining durch Aufzeigen von Optimierungsansätzen, bzw. Ansatzpunkten für eine tiefere Ursachenforschung, bewiesen werden. Diese Ansätze sollten in der zweiten Phase weitergehend untersucht und zur Modellbildung genutzt werden.

Den Ausgangspunkt der Untersuchungen stellte die Hypothese dar, daß es neben bereits untersuchten Einflussfaktoren auf die Filtrierbarkeit, wie Filterkonstruktion, dem Filterhilfsmittel und den Eigenschaften des zu filtrierenden Bieres [z.B. 5, 6, 7] weitere gibt, deren potentielle Relevanz bisher noch erkannt oder nicht ausreichend berücksichtigt wurde und die sich mit Hilfe der Datenauswerteverfahren des Data Mining aus bereits vorhandenen Produkt- und Prozessdaten, d.h. aus chargenspezifischen Laboruntersuchungen und aus den Aufzeichnungen der Prozeßleitsteuerung, ableiten lassen.

Als wichtigstes Werkzeug wurde Data Mining gewählt, das Kernstück des Prozesses „Knowledge Discovery in Databases“ [8]. Es stellt für viele Anwendungen eine hervorragende Möglichkeit dar, aus einer großen Gesamtheit, auch unvollständiger oder verrauschter Daten [9, 10, 11, 12] relevante Zusammenhänge zu extrahieren [13] und sie in einer geeigneten Struktur darzustellen [14].

## **3 FORSCHUNGSZIEL / LÖSUNGSWEG**

### **3.1 Forschungsziel und angestrebte Forschungsergebnisse**

Das Ziel der beantragten Projekte bestand im Aufzeigen eines Weges zur frühzeitigen Abschätzung der Filtrierbarkeit eines Bieres. Des Weiteren sollten mit Hilfe chargenspezifischer, filtrationsrelevanter Prozessgrößen und ihrer Auswirkung auf die Filtration mögliche Störungen im Filtrationsablauf schon im Voraus erkannt werden, um die Parameter der Filtration selbst (z.B. Druckverlauf, Kieselgurdosage) schon frühzeitig abschätzen zu können.

Dazu sollte auf der Basis vorhandener Daten Prozeßwissen extrahiert bzw. solche Faktoren identifiziert werden, die einen signifikanten Einfluß auf den Filtrationsprozeß ausüben. Diese Wissensextraktion sollte mit geeigneten Methoden des Data Mining durchgeführt werden und ein daraus erstelltes Modell die unterschiedliche Testverfahren zur Vorabschätzung des Filtrationsverlaufes [z.B. 15, 16] ablösen, auch wenn erste Ansätze zur Vorhersage der Filtrierbarkeit eines Bieres aus Größen vorgeschalteter Prozeßschritte bereits aufgezeigt werden [17].

Das Vorhaben stützte sich auf die Arbeitshypothese:

Es existieren bisher unbekannte Faktoren, die zu einer Prozeßvorhersage und -optimierung bei der Filtration eingesetzt werden können.

Auf der Identifizierung dieser Faktoren sollten die Beschreibung und Optimierung der Filtration basieren. Das Potential des gewählten Ansatzes sollte zudem auch durch die Erarbeitung von Ansätzen für eine erweiterte Ursachenforschung und einer

tieferegehenden Datenauswertung nachgewiesen werden, um die detektierten Einflüsse zu konkretisieren.

Aufbauend auf der Ausgangssituation, dem Stand der Technik und eigenen Erfahrungen wurden aus der globalen Fragestellung folgende Unterziele unmittelbar abgeleitet:

1. Detektion relevanter Parameter
2. Modellierung der Filtration
3. Strategien zur vertiefenden Ursachenforschung.

Die Ergebnisse der beschriebenen Teilziele lassen sich dabei wie folgt in Kürze zusammenfassen.

#### Zu 1: Detektion relevanter Parameter

Bei diesem Unterziel stand die Entwicklung und Analyse geeigneter Datenpools aus chargenbezogenen Betriebsdaten der mit einer umfassenden Datenerfassung ausgestatteten beteiligten Brauerei im Vordergrund.

Mittels unterschiedlicher Data Mining Verfahren sollten daraus die für die Filtration relevanten Parameter erarbeitet werden. Die Anwendung dieser informationstechnologischen Algorithmen resultierte in der Erstellung rein datenbasierter Modelle, deren Signifikanz sich mit Hilfe der Validierungsdaten und dem Wissen der beteiligten Brauereiexperten überprüfen ließ. Dazu erwies sich die Erstellung unterschiedlicher Datenpools für Trainings- und Validierungsdaten als zwingend notwendig.

#### Zu 2: Modellierung der Filtration

Die Erstellung eines Modells der Filtration stellt die Grundlage ihrer Optimierung und der Prädiktion der Filtrierbarkeit eines Bieres dar. Die Basis dieses Modells bildeten die mit Hilfe des Data Mining gewonnenen Erkenntnisse. Bedingt durch die kurze Vorhabendauer beider Projekte wurden geeignete Ansätze mit dem Ziel der Konkretisierung in einem Folgevorhaben für die Erstellung eines Modells entwickelt. Aufgrund der langjährigen Erfahrung des LFP auf dem Gebiet der kognitiven Algorithmen wurde eine der Alternativen auf der Basis von Fuzzy Logik aufgestellt.

#### Zu 3: Strategien zur vertiefenden Ursachenforschung

Die Präzisierung des Modells und seine konkrete Umsetzung sollen im Rahmen eines bereits beantragten Folgevorhabens durchgeführt werden, das auf den Erkenntnissen der hier beschriebenen Projekte aufbaut. Es konnten jedoch geeignete Strategien aufgezeigt werden, in welcher Weise sich das Potential der unter Punkt 2 gewonnenen Ergebnisse in Brauereien umsetzen lässt.

## **3.2 Ergebnisse**

### **3.2.1 Zusammenfassung**

Die erzielten Ergebnisse demonstrieren weltweit erstmalig die erfolgreiche Anwendung von Data Mining Methoden auf die Wissensextraktion in der Braupraxis. Daraus lässt sich auch die mögliche Übertragbarkeit dieser Methodik zur Lösung von Problemen in anderen Prozessschritten folgern.

Die Prozess- und Chargendaten einer ausgewählten Brauerei wurden auf für die Filtration relevante Einflußfaktoren hin untersucht. Nach intensiven

Voruntersuchungen kamen dabei Cluster- und Entscheidungsbaumverfahren des Open-Source Statistik-Pakets R [18] zur Anwendung. Die im Folgenden ausführlicher diskutierten Ergebnisse können wie folgt zusammengefasst werden:

- mit den am Filter aufgenommenen Prozeßdaten lassen sich Filtrationen differenziert bewerten;
- die durch ein algorithmisches Verfahren gewonnenen Bewertungen stimmen mit der Praxiserfahrung überein;
- es gibt deutliche Hinweise auf einen Einfluß der Lagerzeit auf die Filtrierbarkeit;
- ein Einfluß der Tankgeometrie ist in ersten Untersuchungen sichtbar, ließ sich jedoch aufgrund der nicht ausreichenden Anzahl von Daten mit statistischen Methoden noch nicht als signifikant einordnen;
- um den Modellierungsansatz hinsichtlich seiner Präzision und Erweiterbarkeit zu verbessern, erwies sich die vertiefende Beschreibung vorliegender Daten mittels mathematischer Methoden als unabdingbar;
- die Güte von Filtrationen kann getrennt hinsichtlich der Produktqualität und optimaler Verfahrensgrößen betrachtet werden.

### 3.2.2 Datenvorverarbeitung

Als Grundlage für die Untersuchung der relevanten Einflußfaktoren sollten bereits erfolgte Filtrationen mehrerer Jahre mit informationstechnologischen Verfahren klassifiziert werden. Dazu wurden die von der Brauerei erhaltenen Daten in einem ersten Schritt in das für die angewandten Verfahren notwendige Format überführt, d.h. in Datensätze, bestehend aus den Trübungswerten bei 25° und 90°, Druckanstieg, Kieselgurverbrauch und Volumenstrom. Durch die anfängliche Beschränkung auf die erste Charge am frisch angeschwemmten Filter und die Konzentration auf eine Biersorte wurden zunächst die Vorgeschichte des Filters als auch die Rezeptur als mögliche Einflussgrößen eliminiert.

Als nächster Schritt erfolgte die Bereinigung der Datenbasis durch Löschung von Datensätzen, bei denen einzelne Werte, z.B. Trübungen, fehlten bzw. aus Plausibilitätsgründen nicht berücksichtigt werden sollten, z.B. aufgrund eines defekten Sensors. Damit bestand die Datenbasis aus Werten mit hohem Informationsgehalt. Letztendlich verblieben von anfänglich 1937 Datensätzen noch 39. Diese wurden mittels der Skalenendwerte aus der Prozesssteuerung bzw. im Falle der Labordaten mittels technologisch sinnvoller Werte, auf das Intervall [0, 1] standardisiert. Dieser Schritt stellt sich als erforderlich dar, da die Bedeutung einer Prozeßgröße in einem Auswerteverfahren auch auf ihrem numerischen Wert beruhen kann.

### 3.2.3 Auswahl der Data – Mining Verfahren

Als besonders geeignet für das Erreichen des Forschungszieles erwiesen sich Cluster- und Entscheidungsbaumverfahren, wie sie im Folgenden kurz beschrieben werden.

Clusterverfahren suchen nach Ähnlichkeiten bzw. Unterschieden einzelner Elemente einer Datenmenge und unterteilen diese in möglichst unähnliche Gruppen, deren Elemente sich jedoch in den einzelnen Merkmalen möglichst gleichen, siehe Abb. 1. Die Festlegung der sinnvollen Anzahl von Gruppen trifft der Anwender aufgrund der Abstände zwischen den einzelnen Gruppen und der Zielsetzung der Anwendung.

Da das Verfahren entsprechend dem unter unterschiedlichen Alternativen gewählten Algorithmus vorgeht, erfolgt das Einbringen von fachlichem, im Sinne des bearbeiteten Projektes brauspezifischem, technologischem Wissen, erst mit der Interpretation der Ergebnisse, d.h. mit der Bestimmung der Clusterzahl und der weitergehenden Untersuchung der einzelnen Gruppen und ihrer Elemente.

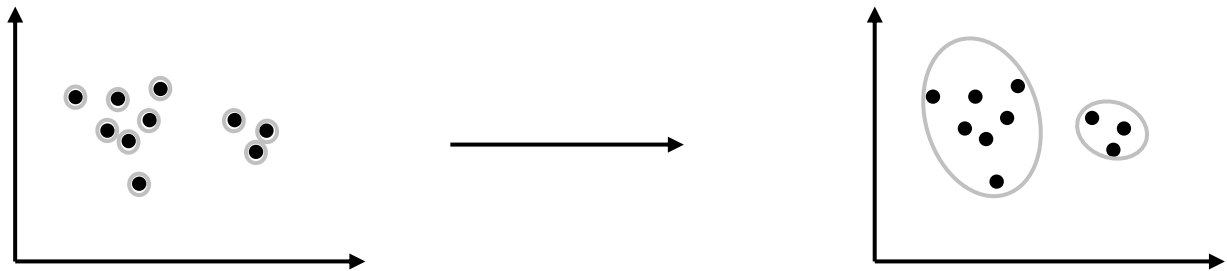


Abb.1: Bildung von zwei Gruppen aus einzelnen Elementen (Datensätzen) mittels hierarchischem Clusterverfahren.

Entscheidungsbaumverfahren benötigen zur Ausführung eine bereits vorhandene Gruppeneinteilung, die z.B. mit Hilfe eines Clusterverfahrens gewonnen werden kann. Sie teilen die zur Verfügung stehenden Daten stufenweise anhand ihrer Merkmale in die vorgegebenen Gruppen ein, wie es in Abb. 2 dargestellt ist.

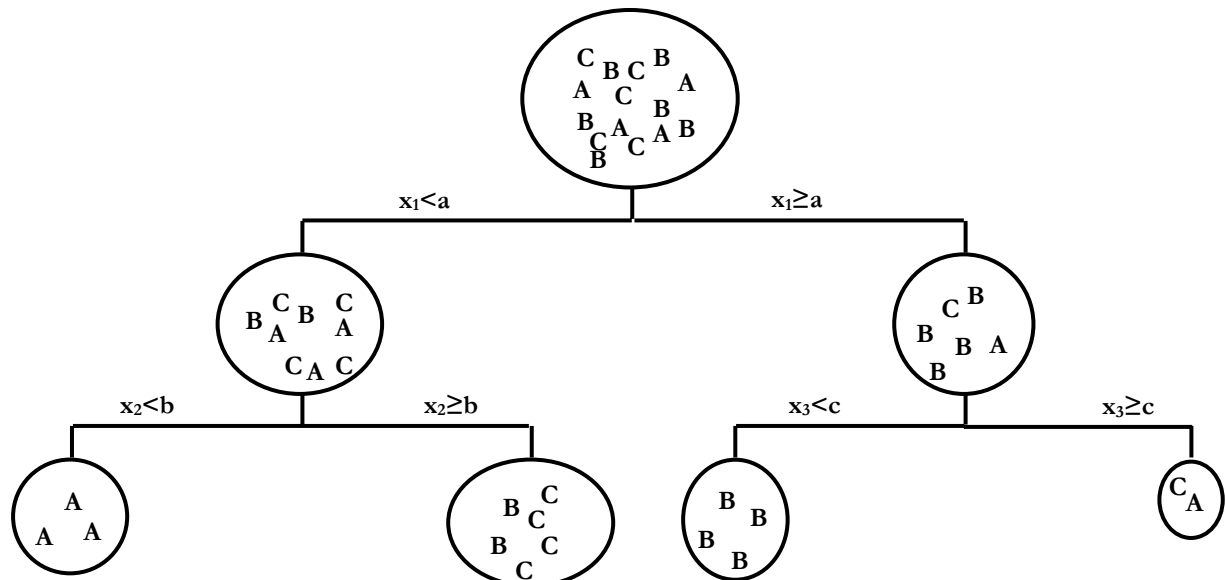


Abb. 2: Vorgehen des Entscheidungsbaumverfahrens. Die Daten wurden durch ein Klassifizierungsverfahren in die Gruppen A, B und C eingeteilt. Die Strukturierung erfolgt anhand der Merkmale  $x_1$ ,  $x_2$  und  $x_3$ , sowie deren Schranken  $a$ ,  $b$  und  $c$ .

Dazu wird ein Merkmal gewählt und die Datenmenge aufgrund einer errechneten Schranke jeweils so in zwei Untergruppen aufgeteilt, daß sich ein vorgegebenes Homogenitätsmaß verbessert. Letztendlich sollen in möglichst wenigen Stufen möglichst homogene Gruppen (hinsichtlich der vorgegebenen Einteilung) entstehen. Mit diesem Verfahren können insofern fachliche Hinweise erhalten werden, als die Auswahl des zur Teilung angewandten Merkmals einen Hinweis auf dessen Bedeutung geben kann. Zusätzlich lassen sich Folgerungen aus dem numerischen Wert des Merkmals ziehen.

### 3.2.4 Klassifizierung der Filtration

Mit dem ausgewählten Clusterverfahren, einem hierarchischen bottom-up Verfahren mit Wards' Algorithmus, wurden die Filtrationen aufgrund ihrer Merkmale, d.h. der

Trübungswerte bei 25° und 90°, dem Druckanstieg, dem Kieselgurverbrauch und dem Volumenstrom, unterschiedlichen Klassen zugeteilt. Die Bewertung einer Klasse im Hinblick auf ihre Merkmale erfolgte durch Vergleich des Gruppenmittelwertes mit dem Mittelwert aller untersuchten Filtrationen und seiner Beurteilung als „besser“ oder „schlechter“. Mit der beschriebenen Methodik ließen sich Filtrationen letztendlich in die fünf Klassen „sehr gut“, „gut“, „mittel“, „schlecht“ und „sehr schlecht“ einteilen.

Zusammen mit den o.a. Merkmalen wurde die so definierte Klasseneinteilung mittels eines Entscheidungsbaumes untersucht. Abb. 3 stellt das Ergebnis in vereinfachter Form dar. Hierbei erwiesen sich bei den Daten der untersuchten Brauerei die Trübungswerte bei 25° als wichtiges Entscheidungskriterium zur Beurteilung der Filtration. Dieses Ergebnis läßt sich einerseits zur geplanten Ursachenforschung heranziehen, andererseits bietet es einen Ansatz zur Erstellung eines Modells.

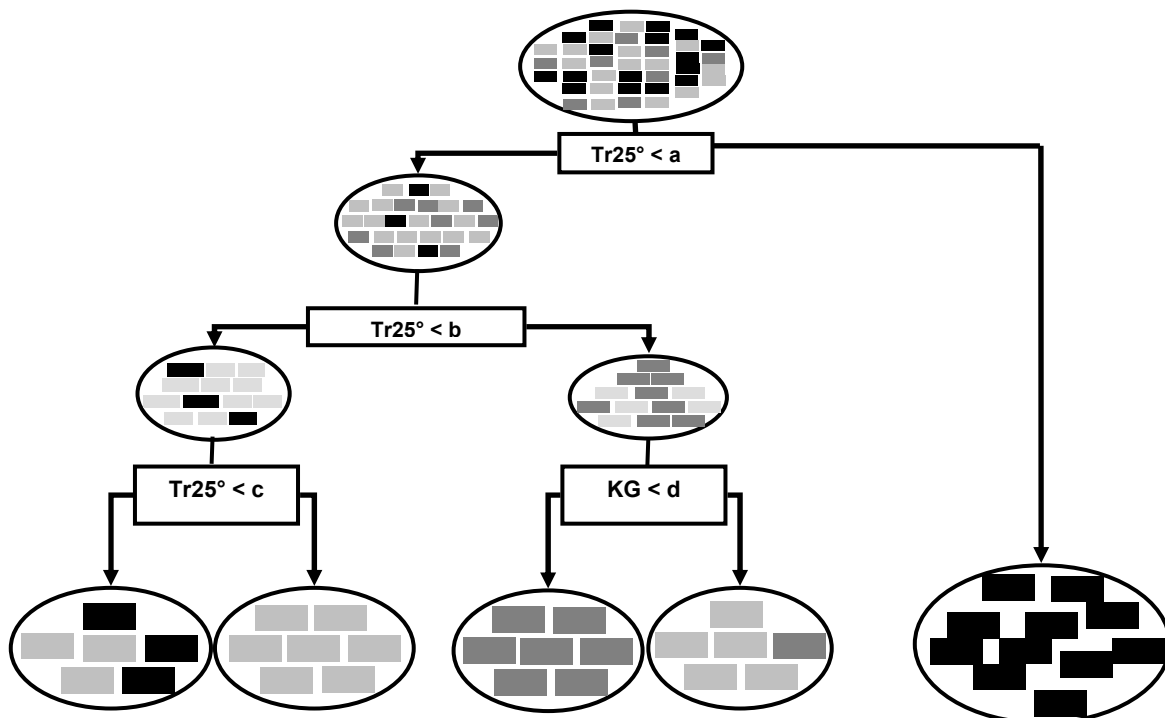


Abb. 3: Entscheidungsbaum zur Filtrationsklassifizierung. Die Buchstaben *a*, *b*, *c* und *d* bezeichnen die Schranken der jeweils genannten Variablen (aus Vertraulichkeitsgründen werden keine Werte angegeben). Die Rechtecke stellen die einzelnen Filtrationen dar, hellgrau: sehr gut + gut, dunkelgrau: mittel, schwarz: schlecht + sehr schlecht.

Ein Vergleich der durch Data Mining gefundenen Grenzwerte der Trübung bei 25° und der entsprechenden Werte der beteiligten Brauerei zeigte eine gute Übereinstimmung der Filtrationsbeurteilung.

### 3.2.5 Filtrationsklassen als Grundlage der Prädiktion

Die gefundenen Filtrationsklassen wurden als Grundlage benutzt, um Labor- und Lagerungsdaten auf Einflüsse auf die Filtration zu untersuchen. Die Ergebnisse deuteten klar darauf hin, daß vor allem der Lagerzeit eine hohe Bedeutung zukommt, aber dass auch ein möglicher Einfluß der Tankgeometrie, d.h. liegend oder ZKT, untersucht werden musste. Insbesondere hinsichtlich der Lagerzeit gibt es eine optimale Dauer, innerhalb der wenige als „schlecht“ beurteilte Filtrationen gefunden werden, wie Abb. 4 zeigt. Dies stimmt sehr gut mit Erfahrungen aus der Braupraxis überein.



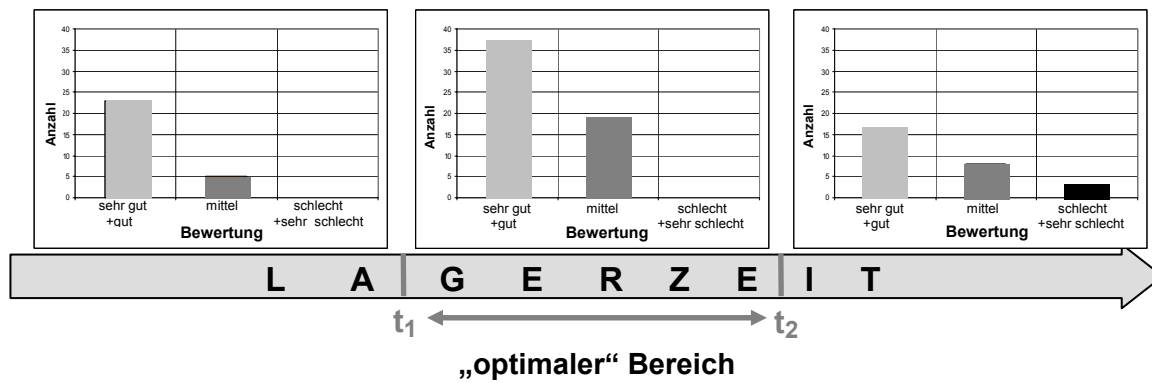


Abb. 4: Auftreten der Filtrationsklassen in Abhängigkeit von der Lagerzeit. Hellgrau: sehr gut + gut, dunkelgrau: mittel, schwarz: schlecht + sehr schlecht.

### 3.2.6 Modellerstellung

#### 3.2.6.1 Allgemeines

Letztendliche Grundlage zur Vorhersage der Filtrierbarkeit eines Bieres bildet die Erstellung eines Modells, das es ermöglichen soll, aus den vorhandenen Daten, hier den Labordaten, bereits im Vorfeld die Filtrationseigenschaften abschätzen zu können.

Dazu kamen drei unterschiedliche Verfahren zum Einsatz:

- Künstliche Neuronale Netze (KNN),
- das Predict – Tree – Verfahren und
- Fuzzy Logik.

Um die Komplexität des Modells in dieser Phase gering zu halten, sollte sich die Vorhersage zunächst auf die Beurteilung „besser“ oder „schlechter“ in Bezug auf eine durchschnittliche Filtration beschränken.

#### 3.2.6.2 Künstliche Neuronale Netze

Künstliche neuronale Netze [19] erlauben die selbständige Generierung einer Beziehung zwischen vorgegebenen Eingangsdaten, hier den Labordaten, und dazugehörigen Ausgangsdaten, hier der entsprechende Filtrationsklasse, durch ein überwachtes Lernverfahren. Abb. 5 gibt den prinzipiellen Aufbau wieder, in dem die Anzahl der Eingangsknoten der Anzahl der im Labor bestimmten Größen entspricht. Die Anzahl der Ausgangsknoten hängt von der Anzahl definierter Filtrierbarkeitsklassen ab.

Nach der Trainingsphase wurden das entwickelte KNN mittels ihm unbekannter Sätze an Labordaten validiert. Mit diesem Ansatz wurde eine Vorhersagbarkeit von 70% erzielt.

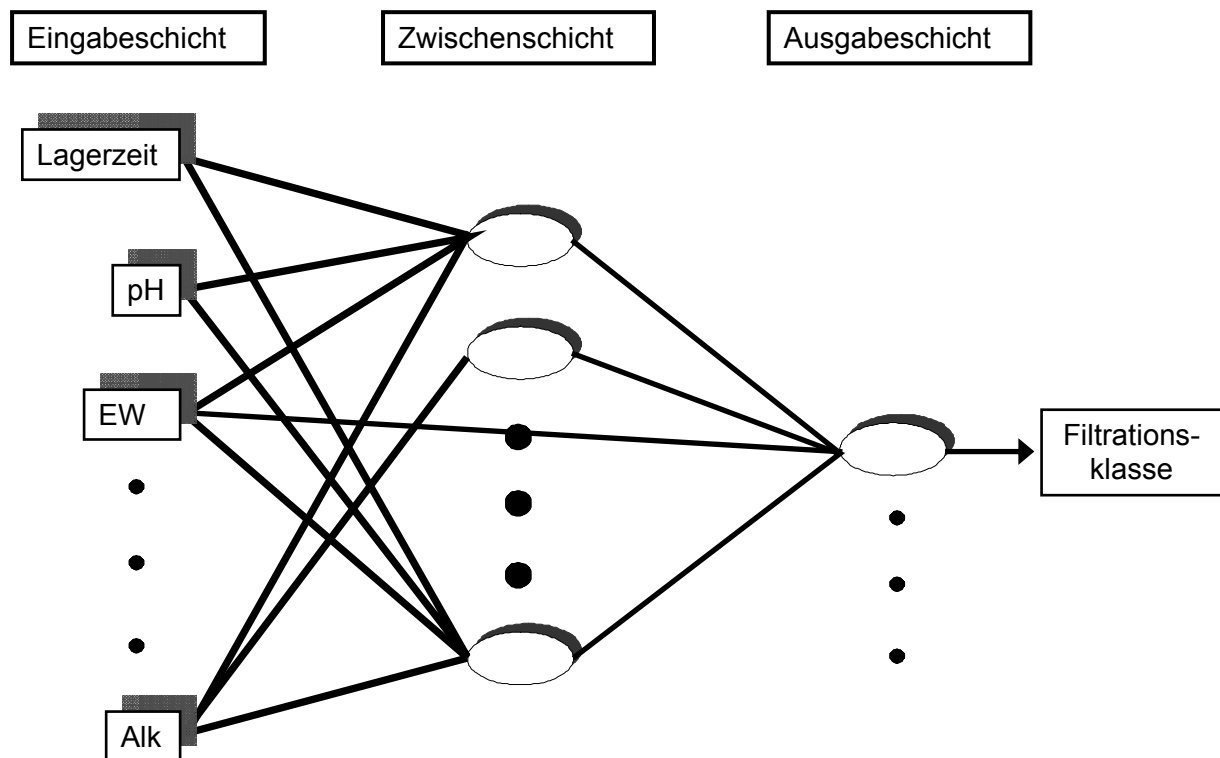


Abb. 5: Prinzipieller Aufbau eines KNN zur Vorhersage der Filtrierbarkeit.

Da es sich bei KNN um sogenannte „black – box – Modelle“ handelt, lassen sich Modifikationen nur in eingeschränktem Maße integrieren, z.B. wenn sich Vorwissen auf geeignete Weise formulieren lässt [20].

### 3.2.6.3 Predict – Tree – Verfahren

Dieser Ansatz beruht auf den bereits beschriebenen Entscheidungsbäumen. Bei einem entwickelten Entscheidungsbaum werden die Prozeßgrößen an den Knoten und deren numerische Werte konstant gehalten. Damit steht fest, auf welchem Weg ein unbekannter Datensatz den Baum durchläuft. Im beschriebenen Vorhaben ließen sich mit dieser Methodik 80% korrekter Vorhersagen erreichen.

Obwohl in beschränktem Maße eine Adaption eines Entscheidungsbaumes durch Anpassung der Grenzwerte eines Knotens möglich erscheint, kann das mit diesem Verfahren erarbeitete Modell nur sehr begrenzt modifiziert werden.

### 3.2.6.4 Fuzzy Logik

Im Gegensatz zu den bisher beschriebenen selbstlernenden Verfahren zeichnet sich der verwendete Fuzzy Logik Ansatz dadurch aus, daß sowohl die Fuzzy Mengen als auch die Regeln zu ihrer Verarbeitung vom Anwender selbst zu erstellen sind. Auch wenn sich dadurch der Aufwand der Modellierung vergrößert, erhöht sich im Gegenzug auch die Flexibilität. Die Entwicklung des Fuzzy Systems beruhte auf den von den Entscheidungsbäumen vorgegeben Grenzwerten der einzelnen Prozeßgrößen. Diese scharfen Grenzen wurden, wie in Abb. 6 dargestellt, in Fuzzy – Mengen umgewandelt.

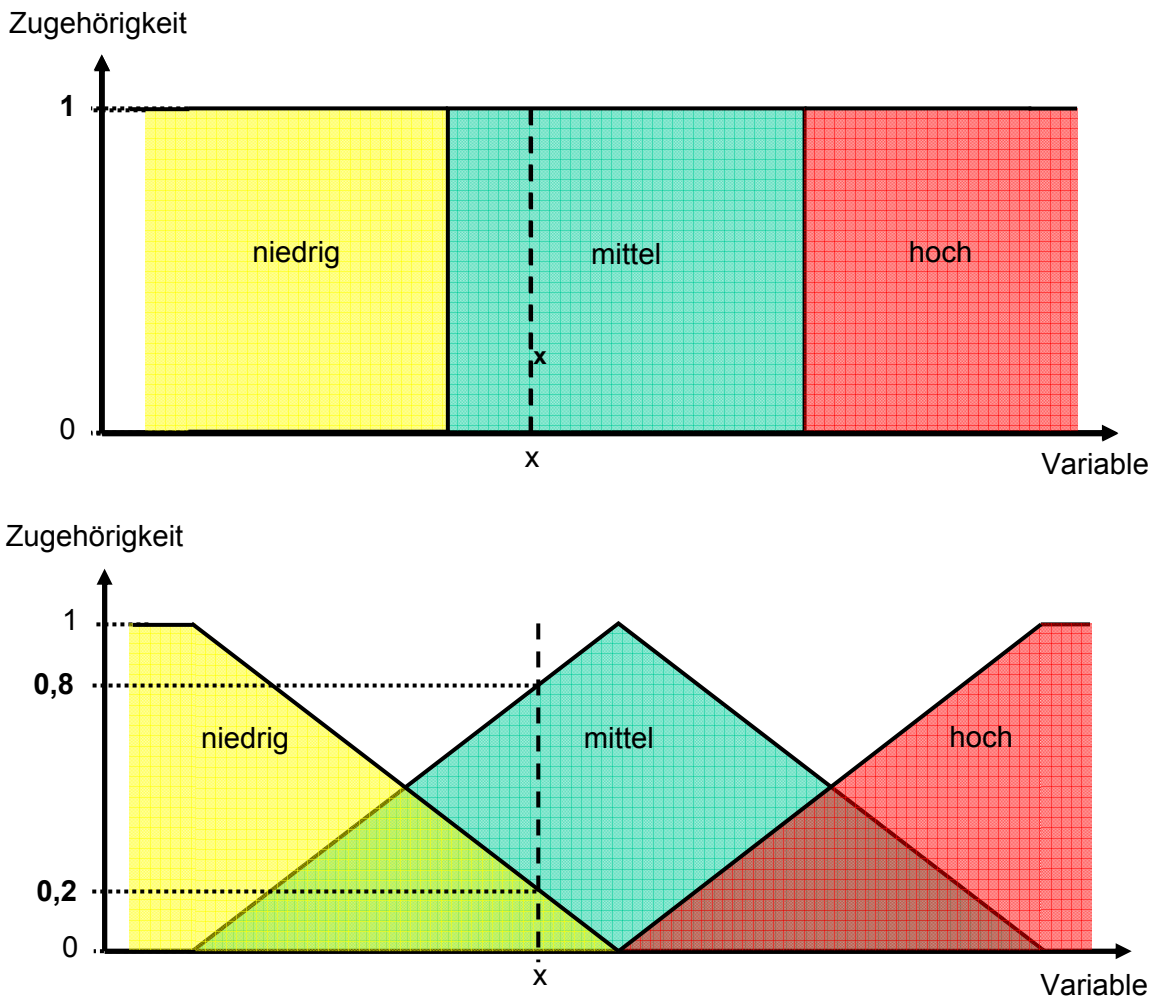


Abb. 6: Die Zugehörigkeit der Variablen „x“ zur Klasse „mittel“ nach dem Entscheidungsbaumverfahren (oben) beträgt „1“, d.h. 100%. Im Fuzzy-System (unten) beträgt die Zugehörigkeit der Variablen x für den gleichen Wert zur Klasse „mittel“ 0,8 (80%) und gleichzeitig 0,2 (20%) zur Klasse „niedrig“.

Linguistische „WENN...DANN...“ – Regeln stellen eine oder mehrere Konsequenzen in Abhängigkeit einzelner oder kombinierter Bedingungen dar.

Das hier erreichte Prädiktionsergebnis von 60% ist vor allem darauf zurückzuführen, daß in der Projektlaufzeit keine Erstellung eines funktionsfähigen Systems vorgesehen war. Dennoch ist in diesem Ansatz das größte Potential für eine zukünftige Weiterführung zu sehen, insbesondere aufgrund der Option, die Erfahrungen von Experten der Brauerei auf vergleichsweise einfache Weise in das Modell einfließen zu lassen.

### 3.2.6.5 Präzisierung der Bewertung

Der gewählte erste Ansatz erwies sich für eine detaillierte Analyse, die zudem den zeitlichen Verlauf der Filtration berücksichtigt, als noch nicht geeignet. Zur Präzisierung des Ansatzes erfolgte eine Übertragung der Filtrationskurven durch Integration der Prozessparameter in monoton wachsende Funktionen, womit sich gleichzeitig eine Glättung der Prozessdaten erreichen lässt. Aus den so erhaltenen transformierten Prozessgrößen können zudem Anlagenkenngrößen ermittelt werden.

In Anlehnung an das Gesetz von Darcy ( $\dot{V} = \frac{\Delta p \cdot A}{\eta \cdot R}$ ) für die ideale Kuchenfiltration kann aus Daten des Prozessleitsystems ein modifizierter Filterwiderstand  $R$

( $R \cong \frac{\Delta p}{V}$ ) berechnet werden, der sich zusätzlich als proportional zur Kieselgurdosage ( $\Sigma KG$ ) erweist, also  $R \cong V_{KG}(T)$ . In ähnlicher Weise konnten die Prozessdaten für Trübung 25° ( $\Sigma Tr_{25}$ ), Trübung 90° ( $\Sigma Tr_{90}$ ) sowie das filtrierte Volumen ( $\Sigma V$ ) dazu genutzt werden, Qualitäts- und Leistungsparametern zu erstellen. Zusätzlich gestattet es diese Modifikation, auch Filtrationen an einem vorbelasteten Filter zu berücksichtigen.

Eine weitere Bewertung der Filtrationen erfolgte durch vordefinierte sogenannte „Dummy-Filtrationen“. Sie erlauben einerseits die Beurteilung der Filtrationen, andererseits aber auch die Erstellung eines Filtrationsbewertungsnetzes, wie es in Abb. 7 dargestellt wird.

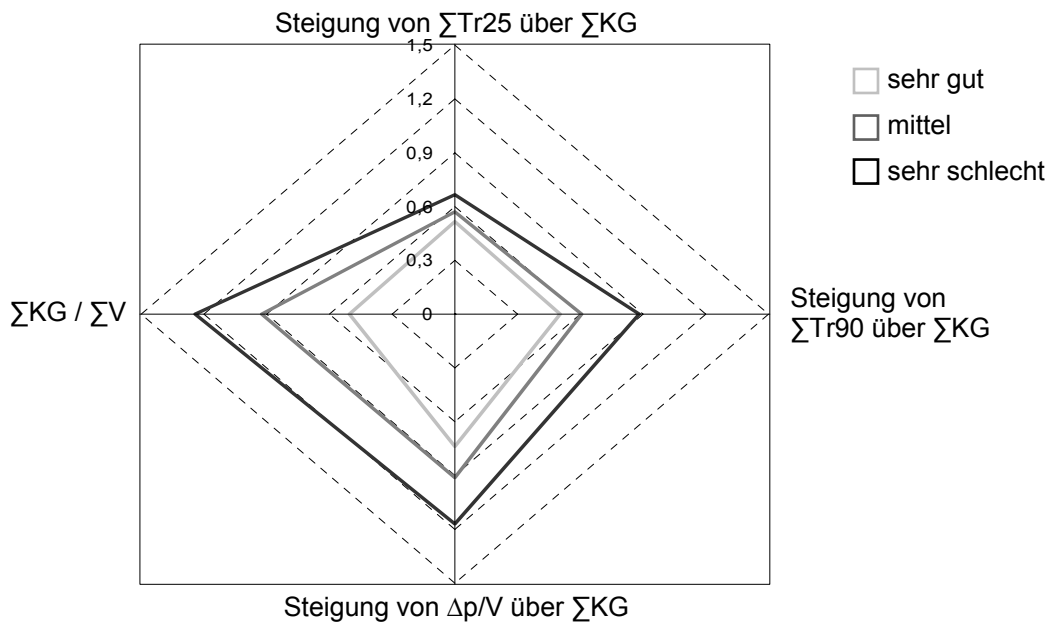


Abb. 7: Filtrationsbewertungsnetz. Die hellgrau, dunkelgrau bzw. schwarz gekennzeichneten Linien geben die Grenzen der Bereiche für sehr gute, mittlere oder sehr schlechte Filtrationen an. Grundlage hierfür sind die Dummy-Filtrationen.

### 3.2.6.6 Untersuchungen zum Einfluss der Reihenfolge der Sorten am Filter

Um Aussagen über die Auswirkungen der Sortenreihenfolge am Filter zu treffen, erfolgte eine Korrelationsanalyse der einzelnen Abfolgen. Erste Ergebnisse dieser Untersuchungen zeigen eine schwache, nicht signifikante Abhängigkeit. Dieses Resultat lässt durch die langjährige Erfahrung der beteiligten Experten erklären, die zu einer Optimierung der Sortenplanung führt.

### 3.2.6.7 Analyse der Labordaten

Zur vertieften Ursachenforschung einerseits und als Grundlage für die beabsichtigte Modellerstellung andererseits erfolgte eine Analyse der vorhandenen Labordaten durch den EM-Algorithmus („expectation maximization“, Erwartungsmaximierung). Mit dieser Rechenvorschrift lassen sich additiv überlagerte Verteilungen von Messwerten trennen und die Parameter der einzelnen Verteilungen angeben.

Am Beispiel des Alkoholgehalts legt Abb. 8 dar, dass sich deutlich erkennbare, mehrgipfelige und durch den EM-Algorithmus separierbare Verteilungen ergeben, womit sich Schwankungen im Prozessablauf oder in der Rohstoffqualität aufdecken lassen.

Die so gewonnene Unterteilung der Daten stimmt mit der des hierarchischen Clusterverfahrens überein. Für die Fortführung des Projektes wird die Eignung der gewonnenen Gruppierung als besonders gut für die Erstellung eines Fuzzy-Systems bewertet.

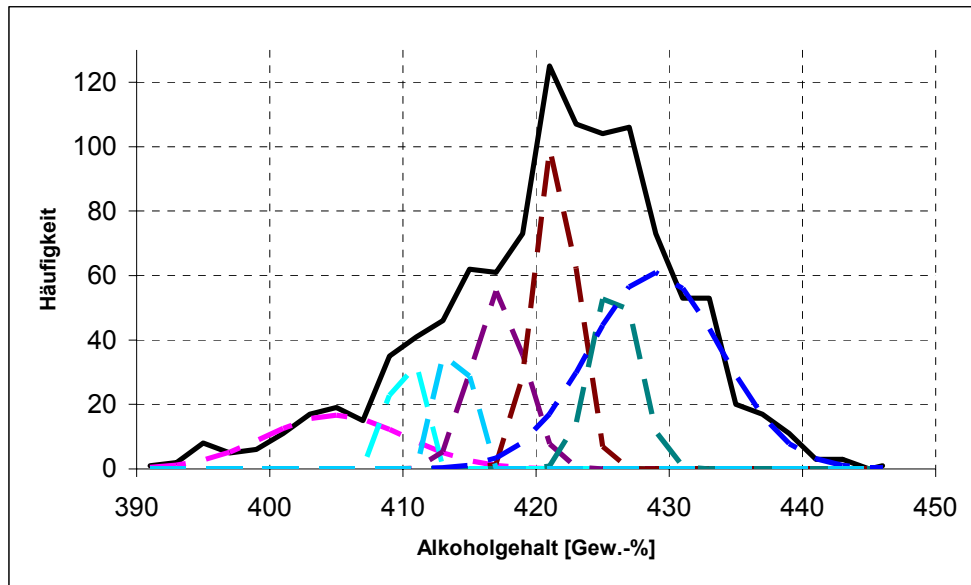


Abb. 8: Darstellung der Häufigkeitsverteilung des Alkoholgehalts (schwarz, durchgezogen) als Überlagerung einzelner Normalverteilungen (unterschiedliche Farben, gestrichelt).

### 3.2.7 Fazit

Die erreichten Vorhersagequoten von bis zu 80% sind im Rahmen der Projektziele des Vorhabens B83, die lediglich in einer ersten Erstellung möglicher Ansätze bestanden, als sehr gut zu beurteilen.

Obwohl das Fuzzy – System in den ersten Untersuchungen die niedrigste Anzahl korrekter Prädiktionen aufwies, wird hierin der erfolgversprechende Ansatz gesehen. Dies begründet sich darin, dass sich hier auf relativ einfache Weise erwünschte Modifikationen realisieren lassen, z.B. die Berücksichtigung von Erfahrungen der am Projekt beteiligten Experten aus der Brauerei. Ein weiterer Vorteil gegenüber den beiden anderen Ansätzen ergibt sich aus der Tatsache, dass fehlende Eingangsdaten, z.B. aufgrund eines fehlerhaften Sensors, nicht notwendigerweise zu einem Versagen oder der Nichtanwendbarkeit des Modells führen, wenn sie durch einen entsprechenden Aufbau der Regelbasis berücksichtigt werden.

Im Folgeprojekt – B83(Forts.) – hatte sich gezeigt, dass die entwickelten Ansätze im Sinne der globalen Zielsetzung, d.h. der Prädiktion der Filtrierbarkeit, eine unzureichende Basis bildeten. Um dem finanziellen Aspekt des Verfahrens Rechnung zu tragen, d.h. weiterhin auf bereits vorhandenes Datenmaterial zurückgreifen zu können, erfolgte eine tiefergehende Analyse der Labordaten. Die daraus gewonnenen Erkenntnisse unterstützen Fuzzy Logik als geeignete Alternative für den Aufbau des „virtuellen Filtrationsassistenten“.

Dennoch kann auch die geeignete Kombination mehrerer Verfahren in einem sogenannten „Hybrid“ als weitere Entwicklung zur Anwendung kommen, in dem sich mehrere Verfahren gegenseitig ergänzen.

## 4 AUSBLICK AUF DAS FOLGEVORHABEN

Die bisherigen Ergebnisse dokumentieren erstmals in eindrucksvoller Form die Eignung sowohl von Data Mining Methoden als auch fortgeschrittener mathematischer Verfahren zur Auswertung der Filtrationsdaten und der Erstellung einer Basis für die Filtrationsvorhersage. Die präzisierte Modellgrundlage lässt die getrennte Filtrationsbeurteilung nach Energieeffizienz und Produktqualität zu und erlaubt dem Anwender damit flexible Prädiktionen mit Rücksicht auf die Unternehmensziele.

In dem beantragten Folgevorhaben soll auf den erarbeitenden Grundlagen aufbauend ein „virtueller Filtrationsmeister“ entwickelt werden, dessen Regelempfehlungen den Filtermeister unterstützen. Neben der Kostenoptimierung soll das System auch auf möglicherweise auftretende Schwierigkeiten frühzeitig hinweisen und so die Möglichkeit zur rechtzeitigen Reaktion zur Verfügung stellen.

## 5 LITERATURLISTE

- 
- 1 Kunze (1998). Technologie der Brauer und Mälzer. 8. Auflage, VLB, Berlin.
  - 2 Hertel, M. (1999). Abfallarme Klärfiltration in Brauereien. *BifA-Text Nr. 12.*, ISSN 0944-5935.
  - 3 Rossmann, A. (2002). Zur Klärung – Gesundheitsgefährdung durch Kieselgur – Entwicklungstendenzen. *Akzente* 3, 18 – 19.
  - 4 Narziß, L. und Esslinger, H.M. (1987). Einflußfaktoren auf die Filtrierbarkeit des Bieres. Teil 1: Filtrationstheorie. *Monatsschrift für Brauwissenschaft* 12, 424 – 427.
  - 5 Esser, K.-D. (1989). Probleme der Bierfiltration. *Brautechnik aktuell*, 177 – 180.
  - 6 Annemüller, D. und Schnieck, T. (1999). A proposal for a filterability and stability check in unfiltered storage beer. *Brauwelt International* 17, 192 – 197.
  - 7 Narziß, L., Reicheneder, E. und Edney, M.J. (1989). Studying beer filtration with an accurate beta-glucan assay. *Monatsschrift für Brauwissenschaft* 7, 277 – 285.
  - 8 Glass, Ä. und Karopka, T. (2002). Genomic Data Explosion – The Challenge for Bioinformatics? *Advances in Data Mining*, 80 – 98, ISBN 3-540-44116-6.
  - 9 Albrecht, D.W., Nicholson, A.E. und Zukerman, I. (1998). Knowledge Acquisition for Goal Prediction in a Multi-user Adventure Game. In: Wu, X., Kotagiri, R. und Korb, K.B. (Ed.). *Research and development in knowledge discovery and data mining*. Proceedings of the 2<sup>nd</sup> Pacific Asia conference PAKDD-98, Melbourne, Australia, 15.-17.04.1998, 1 – 12.
  - 10 Lee, K.C., Park, J.S., Kim, Y.S. und Byun, Y.T. (2000). Missing Value Estimation Based on Dynamic Attribute Selection. In: Terano, T., Liu, H. und Chen, A.L.P. (Ed.). *Knowledge discovery and data mining: current issues and new applications*. Proceedings of the 4<sup>th</sup> Pacific Asia conference PAKDD-2000, Kyoto, Japan, 18.-20.04.2000, 134 – 137.
  - 11 Fayyad, U.M., Djorgovsky, S.G. und Weir, N. (1996). Automating the Analysis and Cataloging of Sky Surveys. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyith, P. und Uthurusamy, R. (Ed.). *Advances in Knowledge Discovery and Data Mining*. AAAI Press, MIT Press, 471 – 493.
  - 12 Garofalakis, M. und Rastogi, R. (2002). Network Data Mining and Analysis. In: Chen, M.-S., Yu, P.S. und Liu, B. (Ed.). *Advances in knowledge discovery and data mining*. Proceedings of the 6<sup>th</sup> Pacific Asia conference PAKDD-2002, Teipei, Taiwan, 6.-8.05.2002, 1 – 12.
  - 13 VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, Gesellschaft für Informatik. *Computational intelligence im industriellen Einsatz*. Tagung Baden-Baden, 11./12. Mai 2000. VDI-Berichte, VDI-Verlag, Düsseldorf, 2000.
  - 14 Witten, I.H. und Frank, E. (2001). Data Mining – Praktische Werkzeuge und Techniken für das maschinelle Lernen. Carl Hanser Verlag, München.

- 15 Stewart, D., Freeman, G. und Evans, E. (2000). Development and Assessment of a Small-Scale Wort Filtration Test for the Prediction of Beer Filtration Efficiency. *Journal of the Institute of Brewing* vol. 106 no. 6, 361 – 366.
- 16 Niemsch, K. und Heinrich, Th. (2000). Raible-Test for Evaluation of Filtration Properties. *Journal of the Institute of Brewing* vol. 106 no. 5, 277 – 285.
- 17 Kreis, S. (2002). Der Einfluß von Polysacchariden aus Malz, Hefe und Bakterien auf die Filtrierbarkeit von Würze und Bier. Dissertation, Technische Universität München.
- 18 R Development Core Team (2003). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>.
- 19 Zell, A. (2000). Simulation neuronaler Netze. Verlag Oldenbourg, München.
- 20 Benning, R.; Becker, T.; Delgado, A. (2001). Initial studies of predicting flow fields with an ANN hybrid. *Advances in Engineering Software* 32, 895 – 901.