Lehrstuhl für Fluidmechanik und Prozessautomation
der Technischen Universität München

Optimisation of Filtration by Application of Data Mining Methods

Michael Eberhard

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. R. Meyer-Pittroff
Prüfer der Dissertation:	1. Univ.-Prof. Dr.-Ing. habil. A. Delgado (Friedrich-Alexander-Universität Erlangen-Nürnberg)
	2. Univ.-Prof. Dr.-Ing. E. Geiger (schriftliche Begutachtung)
	2. Univ.-Prof. Dr.-Ing. K. Sommer (mündliche Prüfung)
	3. Prof. Dr.-Ing. T. Kurz (Technische Universität Berlin)

Die Dissertation wurde am 10.08.2006 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 23.10.2006 angenommen.

Acknowledgements

There are lots of people I would like to thank for a variety of reasons.

Firstly, I would like to thank my supervisor and "Doktorvater" Prof. A. Delgado for his encouragement and advice over all my years at the InformationTechnology Weihenstephan.

Also, I greatfully acknowledge the support of all people at the "InformationTechnology Weihenstephan" and the "Lehrstuhl für Fluidmechanik und Prozessautomation". My special gratitude goes to Dr. H. Petermeier for the extensive support and inspiration.

I wish to thank my examiners, Prof. E. Geiger, Prof. K. Sommer and Prof. T. Kurz for their interest in my work as well as Prof. R. Meyer-Pittroff for taking the chair of examination board.

Furthermore, a thank-you goes to the "Spaten-Franziskaner-Bräu GmbH" for contributing not only data but also lots of experts knowledge and to the "Wissenschaftsförderung der Deutschen Brauwirtschaft e.V." for the endorsement.

Finally, I am indebted to my mother, my sisters and especially Katharina for her endless patience and encouragement when it was most required and for giving me the extra strength, motivation and love necessary to complete this work.

Contents

Chapter 1

Introduction

1.1 Problem Description

Filtration is an essential step in many production processes in food and beverage industry, designed to guarantee product safety and product stability. The history of filtration of beer starts in the middle of the 19th century. At that time Thausing [110] has a very critical view on this new process and recommends it solely for correction of exceptions and microbiological infections. In his opinion only incompetent brewers have to improve the natural clarification of beer in this way. Advantages of this technology, like a shorter time of production and a better clarification of the beer, were obvious. Today consumers expect a product which is free from haze and will not loose this attribute during a longer storage. There are other reasons emphasising the importance of this process like extended distribution channels, elongated shelf life and the stringent specifications on microbiological, sensory and chemical-physical stability [75]. For those reasons the process "filtration" has become more and more important.

The technology of filtration has developed from mass-filter to precoat filter. Today the use of candle filter or horizontal leaf filter is state of the art. Diatomite, the mineral deposit of the eucaryotic earth [102], also known as kieselguhr, is used as filtration aid.

Filterability of beer is an important problem in the brewing process which may fluctuate between or even within batches. An insufficient filterability causes major problems, for example an increasing assignment of personnel and filter-aids. This effects the costs of filtration and, in extreme situations, the breweries ability to supply.

This is also the reason why filtration has become subject of research approaches, but still the precise influence of individual ingredients or of the brewing process technology on filterability is not sufficiently analysed.

1.2 Motivation

Because of the importance of filtration, prediction of filterability was focused by several works [33, 95]. Forecasting was mainly based on malt or unfiltered beer. The problem of predicting filterability according to malt attributes is, that these methods do not consider the technology of brewery. All forecasting systems are based upon laboratory scale experiments. Generally speaking the significance of the described systems is impaired due to a difficult scale-up to production.

Most of the research is concerned with only one particular parameter at a time, focusing on high molecular substances like polysaccharides or proteins, the technology, like the trituration of malt [34], or the geometry of the storage vessels. Mostly single, as independent considered parameters have been targets of the research (SISO: single in, single out).

Recapitulating those investigations, two statements can be made: at first, it is very difficult to study the influence of the technology and the equipments of a brewery, and as a second point, there must be further influences which were not yet determined.

Many breweries use systems for computer-aided production data acquisition (PDA). These systems store a large amount of process data, including knowledge about the process, which is not evaluated so far. Knowledge Discovery in Databases (KDD), also known as Data Mining (DM), includes methods to reveal patterns and coherence in large databases [43]. Analysis and generation of models from data belongs to the domain of classical statistics, but during the last decades Data Mining methods have expanded their areas of application.

These Data Mining methods are successfully implemented in economy, like finance, marketing and telecommunication, and are increasingly used in process technology. The main area of application in technology is sourcing of process problems, modelling and simulation of processes. So far only few efforts have been performed in order to apply Data Mining to processes in breweries yet.

1.3 Target

The KDD process and Data Mining tools provide the opportunity to use the process data, reveal covered information within the data and generate knowledge of the process. Modelling a target attribute by other attributes in the data is perhaps the most traditional Data Mining task [64]. Therefore, the main objective of this thesis is to develop a model for prediction of beer filterability. Fundament for this purpose is formed by the routinely acquired data of the brewing process. Furthermore, this model and the revealed knowledge shall be used to identify factors, which influence the filtration process, and interactions between these factors.

Thus, the model will be a perfect tool for the brewery technician in charge of filtration. As a "virtual filtration assistant" it shall help the technician to control and regulate the filtration by means of kieselguhr dosage and the filtration schedulecan even be adjusted based on the specific needs.

A constraint for implementation is given by the information content available from the used data. As it is one aspect of this thesis to consider only the routinely measured laboratory data, the respective content of information allocate by those data has to be considered.

Finally, the presented work will show the aptitude of the applied Data Mining approach and the used methods for the prediction task. So far a prediction of filtration of about 80% has been accomplished. In addition, this thesis reveals a lack of information of the used data.

Chapter 2

State of the Art

In this work different subjects are addressed. Although no completeness is aimed to, the following description deals with the most important subjects. At first an overview about filtration, particularly the principles of filtration, some aspects of beer filtration and filterability is given, followed by an introduction of Data Mining.

2.1 Filtration

2.1.1 Principles of Filtration

The principle task of filtration is to separate a suspension in its solid and fluid components. Although several principles can be applied for filtration (see for example [86]) the databases available for DM have been collected for a kieselguhr filtration. Thus, in this context, filtration can be considered as a mechanical process which is used for the separation of suspensions with a wide distribution of particle sizes. Compared to other separation processes, filtration stands out with good separation characteristics and low energy demand at the same time [47].

The filter is permeable to the fluid but due to its pore structure the solid phase is not able to pass. The particulate material, depending on its interactions, deposits on the filter and forms a growing porous layer, which takes over the function of the filteraid (cake filtration), or attaches within the pores (deep bed / precoat filtration). Deep bed filtration is applied for the clarification of fluids with a marginal amount of solid particles [65]. Recently, cross flow filtration were at first implemented in brewery industries. Figure

2.1 describes those methods.

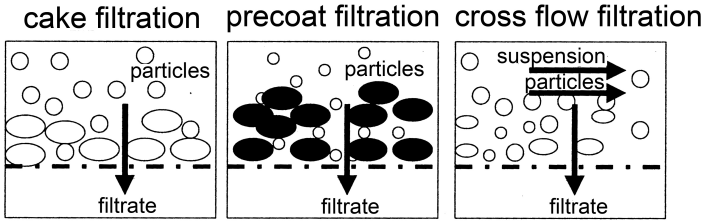


Figure 2.1: Scheme of cake, precoat and cross flow filtration

Filtration systems are subdivided by several criteria. One of these is the size of retained particles. Thereby, four main groups are defined:

- filtration with pore diameter up to 10μm,
- micro filtration with pore diameter up to 1μm,
- ultra filtration with pore diameter up to $10^{-3}\mu$m,
- reverse osmosis.

Table 2.1: Filteraids: fields of application

application	**clarification of**	**filteraid**
food industry	beer, wine, juice, edible oil, treacle, starch hydrolysate	kieselguhr, perlite, silicic acid, wood flour
petroleum industry	petroleum, organic liquids	kieselguhr
metallurgy, mechanical engineering	rolling oil, pickling bath, hardening bath, galvanic bath	kieselguhr

Because of the deposit on or in the filter layer the efficiency of the process fades, either with a loss of pressure difference or, in case of a constant pressure filtration, with a reducing volumetric flow. As the maximum pressure is

limited, a regeneration phase is required when reaching the limit. The use of filteraids provides a longer runtime of the filter. Common filteraids are diatomite, silicic acid and perlite. Their fields of application are described in table 2.1.

At the precoat filtration two layers are established up on the filter with the help of filteraids . The first one, having a coarser structure, forms a supporting layer which is called filter cake. The second layer is applied continously by dosaging kieselguhr to the suspension and therefore growing steadily. Thereby, the available filtration surface is assured, because the continuing addition of kieselguhr inhibits a blocking of surface.

Table 2.2: Diameter of particles in beer

Particle	**Diameter [μm]**
Yeast	5 – 10
Bacteria	0,2 – 2
Tanning protein	0,2 – 1,5
Polyphenol	– 3
β-glucans	10^{-3} – 10^{-2}

Consumer can evaluate the haze of beer, so all particles, which are visible for the human eye, have to be removed. Due to this, the scope of beer filtration is clarification of the product and providing a bright beer during its shelf life. The haze is mainly formed by micro organisms, like yeast cells or bacteria, and high molecular ingredients, like proteins or polysaccharides. Table 2.2 shows some of the relevant substances and their diameters. To assure the brightness of beer until end of shelf life, not only these particles have to be removed, but also those substances which can agglomerate within the time period in question.

To monitor the clarification during filtration, the product is analysed after passing the filter with optical methods. The larger particles, like yeast cells and other micro organisms, reflect applied light, so the reflection is measured in an angle of 25°. The diffusion, caused by smaller particles, like high molecular substances, is measured in an angle of 90° to the light source. The result of this method is presented in form of haze values (h_{25}, h_{90}) and denoted in EBC (European Brewery Convention) units, an unit derived from a formalin haze standard [65]. An overview on this method is given by [54].

With respect to practical relevance, the h_{90} is the more regarded attribute,

because the consumer realizes a turbid beer, caused by many small pending particles, more easily [102].

Even though new filtration methods like cross flow filtration have to be considered, today the common technology to achieve a bright, stable beer is precoat filtration with the filteraid diatomite. Schmidt [102] gives an overview on kieselguhr, its winning, refinement and disposal as well as on its attributes.

Another aspect is the chemical and physical stability of beer. The recent research of Papp [90] and Kusche [71] addresses those problems in detail.

The filtration of beer is influenced by various factors. Schur [103] classifies these influences in four categories:

- filtration plant,
- filteraid,
- mode of operation and
- filterability.

Filtration Plant

Normally, the filtration process is split up into individual steps of defined partial targets. The appropriate adjustment of the subsequent steps is of crucial importance for the filtration success. These steps are commonly separation, precoat filtration, stabilisation and finally a particle filter.

Filteraid

The first common filteraids were asbestos and cotton, today these are kieselguhr (diatomite), perlite or cellulose [86]. Diatomite is purified, dried and grinned skeleton of silicic algae. Diatomite is classified by its permeability and particle size distribution (fine, medium, coarse). The precise combination and amount of these fractions of kieselguhr influences the formation of the filtration cake and the characteristics of filtration. A finer kieselguhr causes a better clarification but effects the permeability of the cake. So the choice and combination of kieselguhr types are a compromise between clarification and pressure increase [65].

Mode of Operation

The level of automation and the control strategies of the plant influence the filtration. Sudden yeast surges, pressure differences or batches with different filterability change the conditions during filtration. Also the demands on shelf life differ between breweries. Modern plants are equipped with haze and pressure meters. Thus, volume flow of beer or the filteraid dosage can be regulated.

Filterability

The filterability has an important influence on filtration. The obtained filtration target should be reached with costs below 0.5 euro/hl. Only sufficient knowledge about the process enables the brewer to react in time. As this is one of the main aspects, filterability is discussed in detail in the following.

2.1.2 Some Aspects of Filterability

After clarification of beer by means of filtration was accepted at the beginning of the last century, it has been noticed that new problems are linked with this process. The common problems are on one side runtime of the filter, because of pressure difference reaching the critical value, and on the other side haze exceeding limit values. Due to this, there are numerous approaches to analyse the process [65].

The term "filterability" has been used in literature many times, but still there is no commonly accepted definition. It is often linked to pressure difference and filtration runtime. In rare cases the clarification has been considered. Kreisz [65] proposes therefore an objective evaluation, which is based on pressure difference, volumetric flow and clarification. Generally, these attributes together with kieselguhr and its characteristics are used to evaluate filtration and its quality.

The basic problem of filterability is to identify the significant upstream influences on this important process type. In literature several approaches try to analyse a variety of parameters. Often these parameters are considered and analysed as independent attributes. On one side influences of ingredients (polysaccharides, proteins, polyphenols, melanoidines, minerals, yeast and bacteria) are mentioned and on the other side several brewing processes from mill to filtration are noted [87, 88, 89]. But not for all of these criteria scientific evidence is shown. The results can be summarised as follows [65]:

- Influences by ingredients:
  - raw materials
    - polysaccharides
    - proteins
    - polyphenoles
    - mineral materials
  - yeast
    - cell count
    - products of autolysis
    - polysaccharides
  - microorganisms
    - polysaccharides
    - mucous substances
- Influences by Technology
  - malt quality
  - milling
  - mashing
  - wort clarification
  - boiling
  - whirlpool sedimentation
  - fermentation
  - storing

Polysaccharides are high molecular carbohydrates. Among these macro molecules, with a molecular weight in a range of several thousands up to some million dalton, the beta- and alpha-glucans are outstanding. Concerning filtration, they are the most discussed beer ingredients. Their appearance in beer depends as well on malt as on technology. Kreisz [65] gives a detailed overview about polysaccharides and their influences on filtration.

2.1.3 Approaches to Prediction of Filterability

There are several systems to predict filterability. Some of these are described in this section. The tests can be divided in two groups: filter tests with constant pressure and tests with constant volume flow. According to Kreisz [65] the following list gives an overview of most important tests:

- test with constant pressure
  - membrane filter test by Esser [33]
  - filtration test by Siebert [105]
  - kieselguhr filter test by Raible [94]
  - kieselguhr filter test by Raible, Heinrich and Niemsch [93]
  - filtration test by Webster and Molzahn [116]
- test with constant volume flow
  - Zuercher test
  - kieselguhr candle filter by Reed [97]

Esser [34] took notice of correlation of filtration in practice and membrane filtration of the same solution. He developed a system based on a filtration of unfiltered beer with a 0,25 μm pores in laboratory scale. The volumes V_1 and V_2 , which have passed the filter at the timestamps t_1 and t_2 are measured and applied to a V/t, t - diagram. The gradient of the connecting line of t_1, V_1 and t_2, V_2 is used as a description of filterability.

In contrast to this, Raible et al. [95] introduced a filtration layer of steel canvas with 15 μm pores. Kieselguhr is precoated on this layer. The filtration time and volume are measured as well as the haze. Based on this volume and runtime the filtration cake coefficient a can be determined as follows:

$$a = \frac{t}{V^2} \left[\frac{s}{m^2}\right] \tag{2.1}$$

This coefficient, Kreisz [43] calls it the filter cake factor, is influenced by the haze of unfiltered beer and the applied kieselguhr. To guarantee the reproducibility the same kieselguhr has to be used. With the cake factor the specific filtration volume can be calculated according to Kreisz [65]:

$$F_{spez} = \sqrt{\frac{3600}{a}} \cdot 0,1 \frac{hl}{m^2 h} \tag{2.2}$$

Another approach, which combines several aspects for a "filterability and stability check" is supplied by Annemueller and Schnick[3].

2.2 Data Mining

2.2.1 Principles of Data Mining

"The convergence of computing and communication has produced a society that feeds on information. Yet most information is in its raw form: data. If *data* is characterised as recorded facts, then *information* is the set of pattern, or expectations, that underline the data. There is a huge amount of information locked up in databases — information that is potentially important but has not yet been discovered or articulated" [119].

In literature Data Mining (DM) is described as the core of "Knowledge Discovery in Databases" (KDD), but also both terms are used synonymic. DM is the process of extracting implicit, previously unknown, and potentially useful information from data. It is also defined as the automatic or (more commonly) semi-automatic process of discovering patterns in data. The discovered patterns must be meaningful in that they lead to some advantage, usually an economical one.

KDD has intersecting aims like statistics, but can be separated from this by several features. The datasets examined by KDD investigators are commonly larger than those of statisticians and statistics is mostly concerned with static data in opposition to the evolving data. But also, DM overlaps in many cases with statistics. Hand [48] discusses the similarities and differences of statistics and KDD.

To be used for the non trivial prediction of new data, the patterns must be expressed. This expression can take place in two different ways, as a black box model or a transparent box model. The transparent box reveals the structure of the pattern, whereas the interior of the black box is effectively incomprehensible. Both modeling techniques yield good prediction, but the results of the transparent box are represented in terms of a structure that can be examined, reasoned about, and used to inform future decisions. These patterns are called structural patterns because they capture the decision

structure in an explicit way [119]. Thus, with transparent box models not only good prediction is realisable, but also gaining knowledge of the examined process.

DM is a multi disciplinary field, that includes database theory, statistics, artificial intelligence (machine learning, pattern recognition) and visualisation methods. Nakhaeizadeh et l. [85] describe the target of DM with a set of sub targets.

Segmentation. Divisioning of data in small, homogeneous, interesting and reasonable subsets or classes. This step is used for preparation of raw data, to create subsets which can be analysed easier.

Classification. Finding common attributes of database objects is the target of classification. These attributes are assigned to defined classes of a classification model. The class names can be predefined or result from segmentation.

Concept description. A concept includes all basic characteristics of a class.

Prognosis. Scope of prognosis is to supplement missing, numeric attributes to objects. In contrast to classification the target variable is a numeric value.

Data description and aggregation. Data description and aggregation help the user to understand data. They allow a description of the basic characteristics of data in compact form. The data aggregation provides a clear structure and helps the user to get an exact image of the targets and data structure.

Anomaly detection. If an object differs from an expected value or norm, then it is an anomaly. This behaviour is significant for a problem which has to be solved; an anomaly can also refer to a yet unknown problem which has to be examined.

Interconnection analysis. The Interconnection analysis searches for models, that describe a correlation between attributes of an object. These interconnections are used to describe the probability, for the appearance of an value with given attributes.

Fayyad et al. [36] also separate the DM-process in sub processes, given by figure 2.2. Chapter three will discuss these processes in detail.

Several methods and algorithms are included under the generic term "Data Mining" [55]. Some of these are Fuzzy Logic [14], Genetic Algorithms [40] or Artificial Neural Networks [77]. Chapter 3 gives an deeper discussion on Data Mining methods in general and the applied ones in detail.

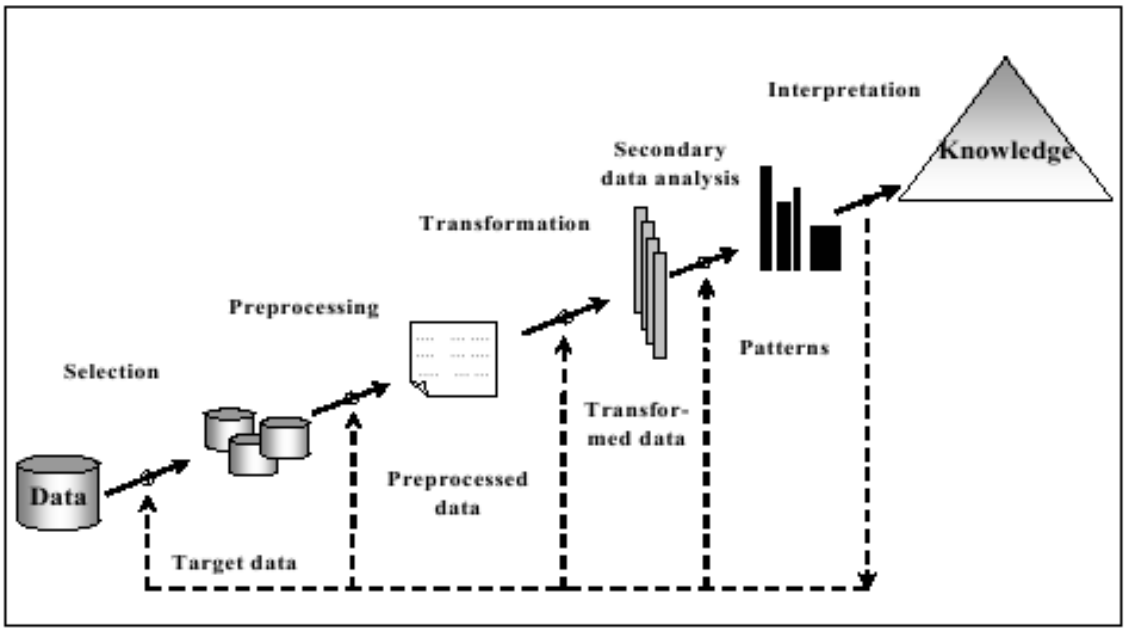


Figure 2.2: Overview of the steps comprising the KDD process

2.2.2 Application of Data Mining Methods

This section gives an overview on fielded applications of Data Mining, according to the papers of Provost et al. [72] and Langley et al. [91] and recapitulating textbooks [62, 122].

Image recognition. Burl et al. [23] and Kubat et al. [68] present applications to image classification, for cataloguing volcanoes on the planet Venus and for detecting oil spills at sea. Fayyad et al. [44] used also a Decision Tree at the Second Palomar Observatory Sky Survey for classifying sky objects like stars or galaxies.

Medicine. Lee et al. [73] and Finn et al. [39] used these techniques in scientific analysis and discovery, for predicting chemical carcinogenicity and for pharmacological discovery.

Chemical process control. Leech [129] predicts the quality of uranium dioxide powder pellets, which are used in nuclear plants, with the help of a Decision Tree method.

Credit decision. Loan companies regularly use questionnaires to collect information about people applying for a credit. Based on 1,014 training cases and 18 descriptive attributes, Michie [82] implements a Decision Tree at American

Express UK.

Marketing. DM methods are often used to analyse market penetration, market development and product development [6].

2.3 Conclusion

The preceding two parts of this section present the state of the art of the two fields Data Mining and Filtration. Together, they provide the motivation and target of this thesis, as they show on the one hand the problems in predicting filtration and on the other hand the aptitude and successfull implementation of Data Mining methods for Knowledge Discovery and prediction tasks. Previous investigations carried out at the Chair of Fluidmechanics and Processautomation of the Technical University of Munich prove those motivation and target. Under the leadership of Professor Delgado the investigation on aptitude of cognitive algorithms started in 1996 [10, 29]. More over, methods like Fuzzy Logic or Artificial Neural Networks are used for recognition of damages on beverage crates [124, 101, 123], modelling of multi-stage high-pressure inactivation of micro organisms [58, 57, 59], state detection and feedback control of anaerobic wastewater treatment [84], modelling and optimisation of fermentations [69, 11, 5, 70] and the prediction of flow fields [13, 12].

Chapter 3

Theoretical Considerations

The first and probably the most prominent sophisticated attempt to define Knowledge Discovery was presented by Fayyad [37] in 1996 ("Knowledge Discovery in Databases is the non-trivial process of identifying valid, novel, potential useful, and ultimately understandable patterns in data."). Since then, several different approaches to describe the KDD process [1, 15, 24, 31, 45, 118] were undertaken. The CRISP-DM (CRoss-Industry Standard Process for Data Mining, compiled by a consortium consisting of DaimlerChrysler, SPSS and NCR, three industrial and commercial organisations concerned with Data Mining and applying Data Mining in their business operations as well as providing services based on Data Mining) initiative started to push forward the standardisation of KDD and DM. The result is a process model for Data Mining that is both a step to a standard methodology as well as an initial guidance in performing Data Mining projects [98]. More detailed information is available at http://www.crisp-dm.org.

This chapter follows the principles of this model. The CRISP view on KDD and DM describes six phases. These steps and their relations are outlined in figure 3.1. Beyond these, relations and interactions between all phases possible exist, moving back and forth is always required. The first phase, business understanding, generally focuses on understanding the business objectives and requirements from a business perspective, then converting this knowledge into a Data Mining problem definition and a preliminary plan designed to achieve the objectives. Section 3.1 adapts this step to the objectives of the presented work. The steps dealing with the data are merged into section 3.2. Here, the used data is introduced, the problems in data acquisition and preprocessing are discussed and the structure of the final dataset is described. CRISPs modelling phase, enfolding selection of modelling tech-

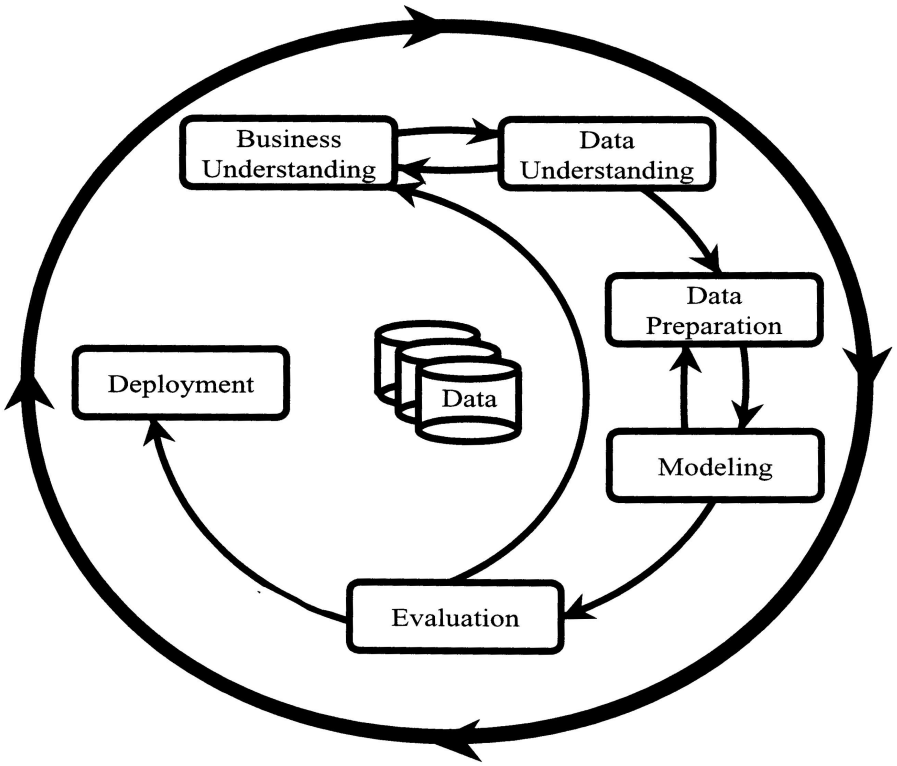


Figure 3.1: The CRISP-DM process model for Data Mining

niques, adaption and calibrating of these methods and the implementation with a discovery system is subject of the sections 3.3, 3.4 and 3.5. Some aspects of the evaluation process are discussed in section 3.6. As mentioned above, this is not a straightforward procedure, but one has to go backwards within the model to obtain the best results and achieve the defined objectives.

3.1 Description of the Task

As defined in chapter 1.3 the objective of this thesis is the optimisation of the filtration process. This is carried out not only in theory but with practical reference. Appendix A gives detailed information on the reference plant of a Bavarian brewery. Due to the very good cooperation with this brewery, a lot of data and experts knowledge was accessible and availed. According

to the CRISP-DM process model, this section deals with the step "Business Understanding". In the presented thesis, this includes getting a first overview on the available data as well as defining the precise targets of DM process.

With respect to problem, motivation and the derived fundamental idea, already described in chapter 1, the main aspect of this optimisation process is the prediction of filterability, with the prediction being an important method of Data Mining. But from this task several subtask and problems arise. Knowledge has to be extracted from already finished filtration, respectively from the archived data, in different steps and ways. These aspects are discussed in the remainder of this section.

3.1.1 Classification of Filtration

To establish a fundament for prediction of filterability the stored data and the information included are needed to be available in an appropriate way. This means that the complexity of the filtration data has to be reduced to a minimum, but still containing the relevant information. Classification is the process of ordering or dissecting complex sets of objects described by high dimensional data into small and manageable units. In the present thesis, this high dimensional data is given by the filtration data stored in the Process Data Acquisition (PDA). The classification of time series like filtration is very complex, regarding all the information contained in these series. Therefore the data is already scaled down before classification. Two approaches of reducing filtration data and classifying filtration are carried out, as described in the following.

First Approach of Classification

In the first approach, the data is reduced to a very minimum. Neither time dependent filtration behaviour nor the state of the filtration plant is taken into account for classification. This alleviates the processing of the data, but also the resulting classification may not suffice the requirements of the task. Another constraint of this approach is given by comparability. With this classification, not all filtration can be operated together, because of differing states of the filter at the beginning of filtration. Alikeness can be approximated only for the first filtration of a batch, due to automated cleaning. Having all this restrictions, this approach shows its eligibility in allowing a first overview on the filtration, possible classes and even in showing possible problems with the data and its processing. First approaches on prediction

of these simple classification can be carried out, but without expecting this prediction to accomplish the main task, paying attention mainly on the first filtration.

Second Approach of Classification

Based on the experiences with the classification of the very simplified filtration data, a more sophisticated classification is aspired. To improve the classification the data is reduced to an optimum, both minimising the amount of data but still offering enough information to describe the filtration. To solve the constraints of the simple classification, generalisation of filtration is aimed at, so that all filtration regardless of their positions within the batches can be analysed together. By this, not only the number of datasets increases but also the significance of the results advances.

3.1.2 Analysis of Input Data

With the above defined task, to predict filtration with the help of Data Mining methods and the available data, this input data has to be analysed. In this case the matter of interest are patterns, associations or structure within the data, which can be used for the prognosis task. The input data considered in this thesis is the laboratory data, the storage conditions and the sequence of filtration. Assuming that this input data and its containing information can be used to predict filtration, the patterns and relationships have to be identified with respect to the corresponding filtration classes.

3.1.3 Prediction of Filtration

This subtask merges the results from the two subtasks described before. Based on the knowledge discovered from the input data a model has to be created to predict filtration classes. Therefore the model has to be constructed or trained with the knowledge of past filtration and the archived data to give a prognosis of filtration using the input data.

3.2 The Data

With respect to the above presented CRISP-DM model, this section deals with the DM steps "Data Understanding" and "Data Handling". It gives an

overview on the available and used data and its processing.

Because of the organisation structure of the reference brewery process data is archived very heterogeneously. Therefore a main concern is to establish a consistent database for the Data Mining tasks. To assure consistency, the used data is limited to filtration data and the laboratory data a priori. This section describes the relevant data and its preprocessing.

3.2.1 Data Structure

As mentioned above, the process data is stored in several places and forms. On one hand the filtration data, based on a PDA, which was implemented together with the filtration plant in 1999, is used in this work. On the other hand the product data, beginning with the laboratory data of the maturated beer just before filtration, is taken in account.

Filtration Data

The filtration data is stored in three parts and by this also in three different data formats. Table 3.1 shows these data, its format and size:

Table 3.1: Data origin and formats

Data	**Format**	**Size**
Trend	binary	1,000 MB
Freie Protokolle	Microsoft Access™	9,500 MB
Schrittprotokolle	Microsoft Dbase™	250 MB

The *Trend* data consist of measured values of 178 sensors, taken during the filtration process, as well as state variables of the filtration process. The values are stored in binary files, with one file per sensor. Each data file, containing the values paired with a reference time stamp, is stored together with an header file containing information about how many values are stored in the data-file. The file names include a time stamp and are separated by calendar weeks. The stored data covers more then 1,000 MB in more then 50,000 files. Relevant values are volumetric flow during filtration, dosage of

kieselguhr, haze, both measured optically at 25° and 90° respectively as well as in- and outlet pressure.

A report on the batch is saved in the *Freie Protokolle*. These reports are managed in a Microsoft Access ™database. The data itself is stored in several (up to 1,500) files, which are evaluated by the database. Among other information, those reports contain details about the preparation of the filter (precoat), the processed beers (including amount) and from which source storage tanks the beer was taken.

The third type are the *Schrittprotokolle*. This data contains information about the process steps, including for example valve and pump states. According to the process recipes the process sequence is recorded.

Laboratory Data

The *laboratory* data is stored in another, independent PDA. The data is available in the DBase™-format. The laboratory data contains information like original gravity, percent alcohol per weight and per volume, both apparent and real extract and attenuation, pH, colour and vicinal dicetones. The analysis is carried out routinely with automated equipment. The results are transfered manually into the laboratory PDA. Thus, wrong and missing laboratory data is not yielded by the analysis itself but by the manual storing.

3.2.2 Data Processing

To meet the requirements of the Data Mining methods the data has to be transferred from those proprietary formats into an more easy accessible format. The CRISP-DM process model includes this process within the step of data preparation. An appropriate solution particular suitable for this situation is a common database. Another concern of preprocessing is data cleaning. It improves the data quality to the required level, to suffice the Data Mining tasks. The following section gives an overview on the database and its structure, the data import and preprocessing.

Database Structure

MySQL database server is the worlds most popular open source database management system and provides a high reliability and security. This database

management system was chosen to store the data.

The data is distributed over several tables. Centre of the database is the table *tbl_cdfp*. This table contains the data from the *freie Protokolle* as shown in Table 3.2. As primary key a field *Filtration_ID* is introduced. It is an unique key for each filtrated beer.

Table 3.2: Data content of database table *tbl_cdfp*

Field	**Type**	**Description**
Filtration_ID	number	primary key, auto increment
auftr_nr	number	derived from production date, containing month and day
charg_nr	number	consecutive numbering of batches starting with 1 each brew year
SZ_Jahr	number	year, 1 or 2 digits
Sorte	text	filtrated beer type
sorte_nr	number	position of filtration within the batch
sorte_hl	number	filtrated volume in *hl*
qt_1_1	number	number of storage department
qt_1_2	number	number of storage tank
⋮	⋮	⋮
up to 15 storage departments and tanks		
⋮	⋮	⋮
qt_15_1	number	number of storage department
qt_15_2	number	number of storage tank

Due to the timestamps, which are not always synchronous and so prohibiting the consolidation into one table, they are stored in one table per each sensor. These tables are labelled tbl_meas*NNN* where *NNN* is the number of the sensor according to the documentation of the PDA (see Appendix for a detailed description). The tables contain only the timestamps paired with the corresponding values (Table 3.3). The digital measured values (state

variables) are stored in the table *tbl_measdigi*. Table 3.4 gives an overview on the relevant fields of this table.

Table 3.3: Data content of database tables *tbl_meas*NNN

field	**type**	**description**
time_int	number	time in seconds, starting at 01.01.1970 00:00
value	number	measured value of corresponding sensor

Table 3.4: Data content of database table *tbl_measdigi*

field	**type**	**description**
time_int	number	time in seconds, starting at 01.01.1970 00:00
b1	boolean	preparation of kieselguhr filter
b2	boolean	processing filter in circuit flow
b3	boolean	filling of first running or feints tank
b4	boolean	filling of pressure tank
b5	boolean	gully
b6	boolean	weighting of kieselguhr
b7	boolean	transferring of kieselguhr

The table *tbl_labor* contains the laboratory data, table 3.5 shows the regular measured and documented attributes.

Data Import

The data import is subdivided in two steps. The *Trend* data, representing the process data of filtration, is converted from the binary files to simple text files with comma separated values (csv-files) of timestamp and sensor value. This is carried out with the help of a self written program using Java™. During the import process of data it is already assured that only consistent

Table 3.5: Data content of database table *tbl_labor*

field	**type**	**description**
Datum_str	text	date of analysis (tt.mm.yyyy hh:mm:ss)
Bereich	text	storage department
Sorte	text	analysed beer type
Nummer	number	composition of storage department and tank, depending on beer type
TK_voll	text	date of filling the tank (tt.mm.yyyy hh:mm:ss)
STW	number	original gravity
ALK_G	number	alcoholic strength (mass)
ALK_V	number	alcoholic strength (volumetric)
ES	number	apparent extract
VGS	number	apparent degree of attenuation
EW	number	real extract
VGW	number	real degree of attenuation
PH	number	pH-value
Farbe	number	colour

time series are imported. Adjacent, the text files are imported into to the MySQL database, resulting in one table per sensor containing the two fields timestamp and value.

The meta information of the filtration, consisting of task number, batch number, year, beer type, position in batch, amount of kieselguhr, begin and end of filtration, filtrated volume and the source tanks and departments, are merged into the database similarly. Therefore the Microsoft Access™database and its scripts (VBA™, Visual Basic for Applications) are modified to combine the data of the original Access database and the *Schrittprotokolle*. The resulting datasets are also stored temporarily in csv-files (comma separated values) and afterwards appended to the adequate table of the MySQL database.

Data Preprocessing

A major problem in the Data Mining process is the handling of dirty data in the data source. Broadly, dirty data include e.g. missing data and wrong data. Dirty data is defined as data which causes the application or rather the user to end up without a result or with a wrong result. The sources for dirty data can be several, including data entry errors by humans or computers, data transmission errors by computer systems or even bugs in the data processing computer system.

Therefore the data must be cleansed to remove or repair dirty data before data analysis algorithms are applied. Kim et al. [60] developed a taxonomy of dirty data, helping to understand the impact of dirty data on Data Mining and on the techniques for dealing with it.

Concerning the data used in this work two forms of dirty data exist which are treated in different ways. These forms are missing data and not missing but wrong data.

- Missing data
  In this work missing data is treated with respect to its impact on the results of the Data Mining algorithms. Especially the time series of filtration process are sensitive to missing data caused by defective sensors or faulty data transmission. Thereby minor missing data, e.g. few values of a time series, can be compensated and have no negative effect on the results respectively, whereas complete missing time series inhibit the analysis. Therefore, they are not taken in account. Also datasets with missing laboratory data are not applied.

- Not missing but wrong data
  This kind of dirty data is hard to identify. It can be caused equally by defective sensors or faulty data transmission. To avoid wrong data, the values are proofed to be in the measuring range of the according sensor. So obviously wrong data is eliminated.

3.3 Discrimination of Methods

The first step in modelling, the adjacent step of CRISPs DM model, is to select the right modelling technique that is to be used. As shown in section 3.1, for accomplishing the defined objective of the present work several different types of modeling techniques are needed. Fig. 3.2 gives an overview on the typical methods used for Data Mining tasks. In this section the typical Data Mining methods are briefly introduced and their aptitude for the tasks of this thesis is discussed. The consequent section gives an more detailed introduction to the applied methods.

As discussed in section 3.1 classification methods are needed as well as methods for prediction or pattern recognition. These types and the available data identify and destine the appropriate methods. Therefore, based on the previous sections, the methods are discussed and narrowed down to the applicable ones. One main fundamental idea is thereby to keep the methods as simple as possible and to approximate this methods to the problem.

3.3.1 Classification and Prediction Methods

A major part of this thesis deals with the classification of data, as this is the fundament for finding patterns, revealing knowledge and finally also prediction tasks. Classification is an ambiguous word. It often means assigning a new object or case to one of an existing set of possible cases. In this work it mostly means finding the classes themselves from a given set of cases. Therefore several classification methods are used for automatic discovery of classes in data. In opposite to the supervised learning, which means generation of class description from labelled examples, these methods are called clustering or unsupervised learning. Automatic classification aims at discovering the "natural" classes in data. These classes reflect coherences which make groups of data looking more similar then others. The underlying causal mechanism can be sample biases in data or can reflect some major new discovery in the domain. The discovered classes can be well known to experts or they can

	Data Type				Data Mining Problem		
Data Mining Methodology	*Labeled Data*	*Unlabeled Data*	*Separate Data Records*	*Time Series Data*	*Predication and Classification*	*Discovery of Data Patterns, Associations, and Structure*	*Recognition of Data Similarities and Differences*
Decision trees	X		X		X	X	X
Association rules		X	X			X	X
Artificial neural networks	X	X	X	X	X		X
Statistical analysis of normal and abnormal data		X	X		X		X
Bayesian data analysis	X	X	X	X	X	X	X
Hidden Markov processes and sequential pattern mining	X	X		X	X		X
Prediction and classification models	X		X	X	X	X	X
Principal components analysis		X	X			X	X
Psychometric methods of latent variable modeling	X	X	X		X	X	X
Scalable clustering		X	X			X	X
Time series similarity and indexing	X	X		X	X		X
Nonlinear time series analysis	X	X		X	X	X	X

Figure 3.2: Typical use of Data Mining methodologies for various Data Mining problems and data types

reveal some important facts, which were previously unknown. The discovery of previously unknown structures occurs most frequently when the data

contains many attributes describing each case, because humans are poor at recognising structure in a large number of dimensions.

In this thesis the classification of unlabeled data presents a very important part, as the past filtration have to be classified and labeled. This labeled data will then be the fundament for prediction. Among the availabel methods, the Decision Tree algorithms and Artificial Neural Networks can be excluded, because these methods require labeled data for supervised learning. The outstanding method is Cluster Analysis. Chapter 3.4.1 describes the underlying principles and algorithms. Sometimes Cluster Analysis is enhanced with the help of special techniques like Bayesian networks, Hidden Markov models or special algorithms for clustering time series. In the present case, these additional extensions would exceed the requirements of the task and increase complexity of the model unnecessarily. Warren Liao [114] gives an overview on clustering time series, showing that clustering of raw data implies several problems, e.g. how to deal with different length of the series or divergent sampled data. None of the papers discussed in that survey handles multivariate time series data with different length, as given with this thesis.

For prediction Decision Trees fullfill the requirements of the task as well as Artificial Neural Networks. Another method not always connected to Data Mining is Fuzzy Logic. With its ability of including experts knowledge and rules defined by other (Pattern Recognition) methods, Fuzzy Logic is taken into account for prediction. Bayesian networks and Hidden Markov methods are excluded for the same reasons of complexity.

3.3.2 Pattern Recognition

Discovery of important structure is usually a process of finding similarities, interpreting results, transforming / augmenting data and repeating the circle. This process also needs the interaction of the discovering program and the expert. On the one hand the structure searching program has the ability to search huge amounts of data looking for multi-dimensional structure with outstanding speed and accuracy. On the other hand the expert has domain knowledge that the program lacks.

Among the algorithm for the pattern recognition tasks Decision Trees and Association Rules are the outstanding ones. In this approach Association Rules are not used. Association Rules were developed originally for market basket analysis and so its qualities are fit to this scenario, to find relationships or affinities between unlabeled datasets. But as defined in chapter 3.1.2 it is very important to reveal patterns regarding the filtration classes, the

labeled data. Additionally the Expectation Maximisation (EM) algorithm as described in the following section is used for extracting knowledge from data.

3.4 Methods

This chapter gives an overview on the methods used in this thesis.

3.4.1 Cluster Analysis

Clustering is used to reveal structures in data that can be useful for the analyst by gaining an insight into some structure inherent in the data set. It is one of the most widely used techniques in Data Mining. The aim of Cluster Analysis is partitioning a d-dimensional data set of n-entities or points into k subsets (clusters), so that the data points within a cluster are more similar to each other than to those in other clusters [8]. Some 40 years ago, biologists and social scientists began to develop algorithms to find groups in their data, advanced by computers becoming available. Up to now, clustering methods and algorithms are applied in many domains, including artificial intelligence, pattern recognition, chemometrics, geoscience, marketing, medical research and political science. Kaufmann [56] gives an overview on this development.

Benchmarking and evaluating of cluster algorithms and implementation are general problems. One main problem is measuring the quality of resulting clusters depending on the application problem [44]. Zaït and Messatfa [129] describe this problem in detail. Michaud [82] compares some cluster techniques, whereas Bock [18] is the most important reference for entrenched cluster techniques.

The choice of a clustering algorithm depends both on the data, e.g. type and availability, and on the aspired objectives. The determination of the appropriate method to use is sometimes not easy, especially when a priori arguments may not suffice to narrow down the choice to a single method. Also because of the descriptive and explorative complexion of Cluster Analysis, it is permissible to apply several algorithms on the same data [56].

Cluster algorithms can be characterised by three attributes. These are classification type, partition criterion and construction method. These characteristics are discussed in the following.

Classification type

The different classification types are

- overlapping classes,
- partitions,
- quasi-hierarchical and
- hierarchical structures.

Each of these forms can be exhaustive, so that each element belongs to at least one cluster, or nonexhaustive, having elements without assignment to a cluster. The latter is advisable for datasets with extreme outliers, avoiding heterogenous classes.

Overlapping classes (see Fig. 3.3a) identify a classification having some elements belonging to more then one cluster, without clusters being part of another cluster. In contrast to this, partitioning methods (see Fig. 3.3b) yield classification with objects belonging to no more then one class.

The quasi-hierarchical classification (Fig. 3.3c) is generated by a sequence of overlappings. On each level of the hierarchy the classes have the characteristics of overlapping classes. Comparing the levels, classes from one level are stringently part of a class from the next level. The hierarchical structure is very analog to the quasi-hierarchical, having partions instead of overlappings within the different levels. These structures are characterised with the help of a pedigree, having the coarsest overlapping on the bottom level and the finest on the top. Especially the hierarchical classification (Fig. 3.3d) is often described with a dendogram. These diagrams are constituted top (coarse) to bottom (fine), having, in addition to the pedigree, a distance scale. Figure 3.4 shows a pedigree (a) and a dendogram (b) of the hierarchical clustering mentioned above.

Partition criterion

As it is the goal to find clusterings which both satisfy homogeneity within each subset as well as much heterogeneity between the clusters, a criterium has to model these conditions. The following section gives an overview on several criteria for defining a measure of adequacy or inadequacy for a given partition.

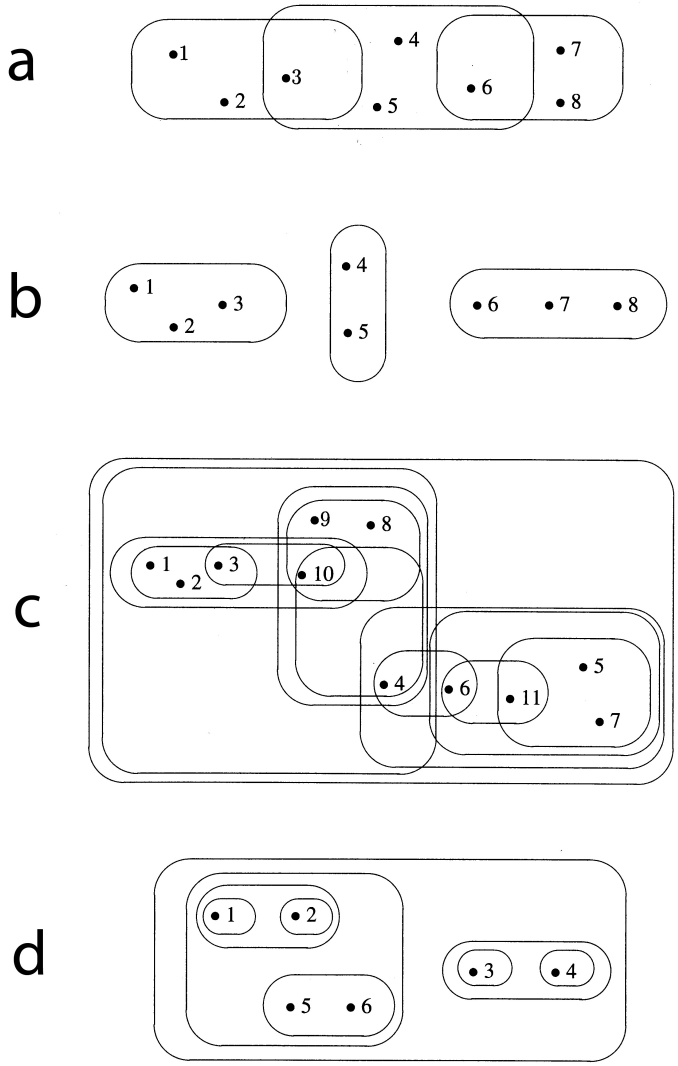


Figure 3.3: Classification types: a) overlapping, b) partitioning, c) quasi-hierarchical and d) hierarchical clusters [56]

Similarity and Dissimilarity of elements and groups of elements can be defined in several ways. Depending on the defining function, certain aspects are emphasised or disregarded. Similarity can be defined as a real-valued function, yielding a real number $s = s(e_i, e_j) = s_{ij}$ for two elements $e_i, e_j \in E$, the universe. This function s can be constricted to the interval $s_0 \leq s \leq s_1$,

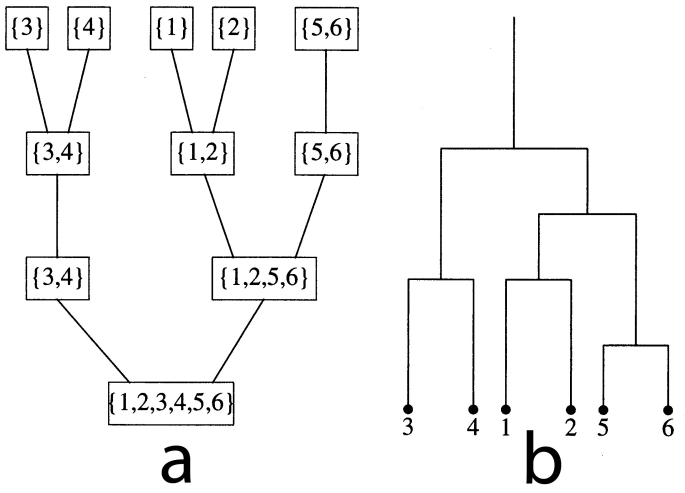


Figure 3.4: Pedigree (a) and dendogram (b), according to the hierarchical cluster seen in fig. 3.3d [56]

with $s_0 = 0$ and $s_1 = 1$, so that $s_{ij} = s_1$ shows the maximum and $s_{ij} = s_0$ the minimum similarity. The following axioms are valid:

$$s_{ij} \leq s_1 \tag{3.1}$$

$$s_{ij} = s_{ji} \tag{3.2}$$

$$s_{ii} = s_1 \tag{3.3}$$

The (n, n)-matrix $\mathbf{S} = (\mathbf{s_{ij}})$ is called similarity matrix. According to (3.2) and (3.3) the matrix is symmetric and all diagonal elements are identical.

Equally, the dissimilarity or distance of elements can be defined as a real-valued function, describing to elements $e_i, e_j \in E$ with a non-negative real number $d = d(e_i, e_j) = d_{ij}$ and having analogue axioms.

$$d_{ij} \geq 0 \tag{3.4}$$

$$d_{ij} = d_{ji} \tag{3.5}$$

$$d_{ii} = 0 \tag{3.6}$$

The (n, n)-matrix $\mathbf{D} = (\mathbf{d_{ij}})$ is called distance matrix. According to (3.5) and (3.6) the matrix is symmetric and all diagonal elements are identical.

Obviously, a major similarity between two elements is connected with a minor distance. Therefore, several functions can be applied to transform similarity functions to distance functions. Examples for these transformations are:

$$s = \frac{1}{(1 + ad^2)} \tag{3.7}$$

$$s = 1 - \frac{d}{a}, \qquad a = \max_{i \neq j} d_{ij} \tag{3.8}$$

$$s = \frac{(H - d^2)}{(H + d^2)} \qquad \text{with} \qquad H := \text{ medial distance.} \tag{3.9}$$

With $0 \leq s \leq 1$ the following transformations are valid:

$$d = 1 - s \tag{3.10}$$

$$d = \frac{1}{2}(1 - s) \tag{3.11}$$

$$d = -\log s. \tag{3.12}$$

Based on this functions and the distance matrix, classifications can be evaluated by concerning their homogeneity and the heterogeneity between the classes. The following functions show some possibilities of describing the homogeneity of classes, the larger the value $g(C_l) \geq 0$ for the class C_l, the more inhomogeneous is the class.

1.: Deviated sum of distances of objects in C_l

$$g_1(C_l) = \frac{1}{c} \sum_{\substack{i<j \\ i,j \in C_l}} (d_{ij}) \tag{3.13}$$

with $c = n_l$ (number of objects) or $c = \frac{(n_l - 1)}{2}$ (number of object pairs). The latter leads to a better evaluation of classes with many objects.

2.: Distance of the two most dissimilar objects in C_l

$$g_2(C_l) = \max d_{ij} \tag{3.14}$$

3.: Distance of the two most similar objects in C_l

$$g_3(C_l) = \min d_{ij} \tag{3.15}$$

4.: Sum of variances

$$g4(C_l) = \sum_{j=1}^{m} s_j^2(C_l) \quad \text{with} \quad s_j^2(C_l) = \frac{1}{n_l - 1} \sum_{i=1}^{n_l} (x_{ij} - \bar{x}_j)^2. \tag{3.16}$$

The heterogeneity of two classes $v(C_l, C_{l*})$ can be described in similar ways. The value v, the degree of heterogeneity, increases with inhomogeneity of the compared two classes.
1. Distance of the two most dissimilar objects in C_l and C_{l*} (complete linkage, furthest neighbour)

$$v_1(C_l, C_{l*}) = \max_{i \in C_l; j \in C_{l*}} d_{ij} \tag{3.17}$$

2.: Distance of the two most similar objects in C_l and C_{l*} (single linkage, nearest neighbour)

$$v_2(C_l, C_{l*}) = \min_{i \in C_l; j \in C_{l*}} d_{ij} \tag{3.18}$$

3.: Medial distance of all object pairs in C_l and C_{l*} (average linkage)

$$v_3(C_l, C_{l*}) = \frac{1}{n_l n_{l*}} \sum_{i=1}^{n_l} \sum_{j=1}^{n_{l*}} d_{ij} \tag{3.19}$$

4. Quadratic, euclidian distance of centre of gravity of the classes (centroid)

$$v_4(C_l, C_{l*}) = \|\bar{x}_l - \bar{x}_{l*}\|^2 \quad \text{with} \quad \bar{x}_l = \frac{1}{n_l} \left(\sum_{i=1}^{n_l} x_{i1}, \ldots, \sum_{i=1}^{n_l} x_{im} \right)' \tag{3.20}$$

To evaluate the heterogeneity of overlapping clusterings, the degrees mentioned above (3.17-3.20) can be applied, after elimination of double entries from the corresponding classes. Also to determine the heterogeneity of hierarchical and quasi-hierarchical clusters the equations are used, provided that the logical difference

$$C_{l*}^* = c_{l*} \setminus C_l \tag{3.21}$$

is used if C_{l*} contains C_l and than $v(C_l, C^*_{l*})$ is calculated.

The overall quality of a Classification C, having g partitions, can be described as the sum of class

$$q_1(C) = \sum_{l=1}^{g} g(C_l) \tag{3.22}$$

homogeneities or as the reciprocal sum of heterogeneities

$$q_2(C) = \frac{0,5g(g-1)}{\underbrace{\sum_{C_l \in C} \sum_{C_{l*} \in C}}_{l<l*} v(C_l, C_{l*})} \tag{3.23}$$

or a combination of both

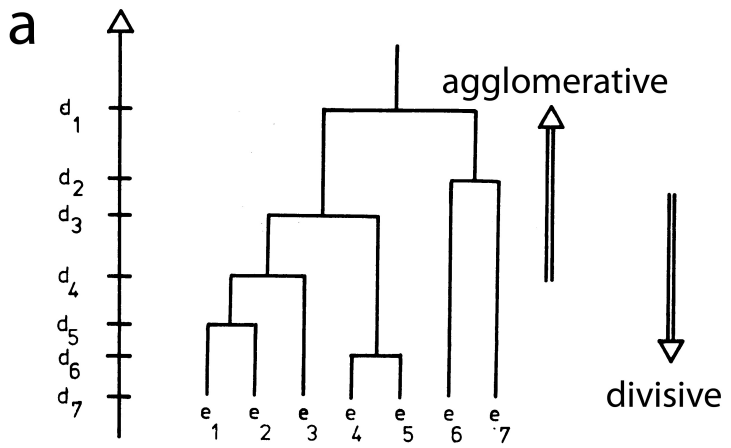
$$q_3(C) = q_1(C)q_2(C). \tag{3.24}$$

Construction method

Due to the impracticability of investigating all possible partitions, clustering methods use various heuristic strategies for obtaining an optimal solution [82]. Traditionally, two different classes of algorithms are used for clustering. On the one hand, these are the hierarchical cluster algorithms which result not in a single clustering but in a refining sequence of clusterings. On the other hand, class of methods are the partitioning cluster algorithms, requiring the number k of classes defined in advance. Commonly, the resulting clusters are visualised in a tree-like structure. Beside these widely used classes of algorithms some other algorithms can be found in literature, mostly used in special, unique applications or with restrictions [44].

Hierarchical clustering is known to be the oldest heuristic approach to obtain a nearly optimal solution [82]. The agglomerative hierarchical clustering starts with n clusters, each having only one element. Iteratively, a pair of clusters is merged into one, so that the number of clusters decreases by one. Which pair of clusters or elements to merge is determined by the best objective function value, obtained by the merge. The iterative merging process stops when only one cluster is left. In contrast to this agglomerative methods stands the divisive hierarchical clustering, starting from one entire set of all

elements (one cluster) iteratively splitting this into two subsets, until each element has its own cluster (fig. 3.5).



b

d_1	{1, 2, 3, 4, 5, 6, 7}	
d_2	{1,2,3,4,5}	{6, 7}
d_3	{1,2,3,4,5}	{6} {7}
d_4	{1,2,3} {4,5}	{6} {7}
d_5	{1,2} {3} {4,5}	{6} {7}
d_6	{1} {2} {3} {4,5}	{6} {7}
d_7	{1} {2} {3} {4} {5}	{6} {7}

Figure 3.5: Agglomerative and divisive clusterings; a) construction of clusters, b) cluster at different levels of distance d_i

A constraint of these methods is given by the pairwise grouping. Once defined clusters can not be divided, whereas non-hierarchical methods can. These partitioning methods separate the given elements simultaneous in k clusters. It is important to note that k is given in advance. Indeed, the algorithms will construct a partition with as many clusters as necessary. Of course, not all values of k are reasonable, so its advisable to apply different values of k. The common idea of partitioning methods is to refine the quality of classification with respect to a partition criterion, starting from an initial partition and rearranging elements iteratively. Mostly, the methods differ by the applied criterion, with each method yielding different classifications, with respect to the extremum of the partition criterion. Obviously, a global

extremum of quality exists, because the number of arrangements is finite, but the approach of an overall iteration of all criterions is not practicable due to the fast increasing number of arrangements. So none of the methods ensures the global extremum to be found.

The following hierarchical and partitioning algorithms are used in this work.

- hierarchical
  - AGNES - agglomerative nesting
    The agnes algorithm constructs a hierarchical cluster by combining the two nearest clusters or elements at each step, yielding as well a dendogram and a banner plot. Also it determines the agglomerative coefficient (ac) which measures the amount of clustering structure. The ac can be used to compare two clustering approaches, but because it depends on the number of observations it should only be used for the comparison of identical datasets, e.g. to evaluate the appropriate partition criterion [50, 108]. Table 3.6 gives an overview of the clustering methods which can be applied for the agglomerative approach.
  - DIANA - divisive analysis clustering
    This method computes a divisive hierarchy of the dataset. Similar to agnes, it provides a dendogram, a banner plot and a divisive coefficient. The algorithm constructs a hierarchy starting with one large cluster containing all elements. Clusters are divided until each cluster contains only one element. At each iteration step the cluster with the largest dissimilarity is selected. To divide this cluster the most disparate element is determined. This element forms the so called "splinter group". Afterwards, the algorithm reassigns elements that are closer to the "splinter group" than to the "old party", yielding two new clusters [50].
- partitioning
  - PAM - partitioning around medoids
    The algorithm aims at finding k representative objects or medoids among the dataset, which describe the structure of the data. The initial k medoids can be given or will be chosen by the algorithm. Based on these medoids k clusters are constructed, assigning each element to the nearest medoid. The algorithm iterates the medoids to find the k representative elements which minimise the sum of dissimilarities [50].

- CLARA - clustering large applications
  Clara is very similar to pam. It was developed to handle large datasets. Therefore, clara does not store all dissimilarities in central memory of the computer. Beside that, the algorithm is identical to pam [50].

Another algorithm used is the k-means algorithm. The k-means algorithm is similar to PAM. The given data is clustered into k groups such that the sum of squares from elements to be assigned to the cluster centres is minimised. These centres can be pre-defined or otherwise they will be chosen by the algorithm. The packages available for the software suite "R" provide several different implementation of the algorithm from which the Hartigan-Wong algorithm is used [50].

3.4.2 Decision Trees

Decision Tree learning is a common method used in Data Mining. A Decision Tree describes a tree structure wherein leaves represent classifications and branches represent conjunctions of features that lead to those classifications [81]. A Decision Tree is also a predictive model [4]. A very good overview on this topic is given by Buntine [22].

A Decision Tree can be set up by recursive partitioning of a dataset into subsets. This process is repeated on each derived subset. The recursion is completed when splitting is either non-feasible or a singular classification can be applied to each element. To avoid large tree structures, the tree may be pruned.

The two most often used algorithms are C4.5 and CART (Classification and Regression Trees). CART was published in 1984 by Breiman [20]. The algorithm constructs binary trees, such that each node has two branches. C4.5 is the successor of the ID3-algorithm firstly introduced by Quinlan in 1986 [92]. Compared to CART, C4.5 is not limited to binary trees. Therefore it can construct more compact trees.

The Decision Tree method is a white box model, that can be understood and interpreted with few effort. The method is very robust and performs well with large datasets. Both nominal and categorical data are handled by the algorithm, data preparation is needed only in basics.

In this work, the CART-algorithm is chosen. It constructs a binary tree using a given response (left-hand-side) to divide the dataset (right-hand-side)

Table 3.6: Agglomerative methods

Method	**Distance**	**Characterisation**
Single linkage (Nearest neighbour)	$\min(d_{pi}, d_{qi})$	contractive, monotone, invariant under monotone transformations of distance matrix, chaining
Complete linkage (Furthest Neighbour)	$\max(d_{pi}, d_{qi})$	divergent, monotone, invariant under monotone transformations of distance matrix, tends to small groups
Average linkage	$\frac{1}{2}(d_{pi} + d_{qi})$	conservative, monotone
Average linkage (weighted)	$\frac{1}{n}(n_p d_{pi} + n_q d_{qi})$	conservative, monotone, weighted by count of elements
Median	$\frac{1}{2}(d_{pi} + d_{qi}) - \frac{1}{4}d_{pq}$	conservative, non-monotone (inversion), only for squared euclidian distances
Centroid	$\frac{1}{n}(n_p d_{pi} + n_q d_{qi}) - \frac{n_p n_q}{n^2} d_{pq}$	conservative, non-monotone (inversion), only for squared euclidian distances
Ward	$\frac{1}{n+n_i}((n_i + n_p)d_{pi} + (n_i + n_q)d_{qi} - n_i d_{pq})$	conservative, monotone, minimisation of variance criterium for squared euclidian distances
Flexible strategy	$\alpha(d_{pi} + d_{qi}) + (1 - 2\alpha)d_{pq}$	$\alpha \to 0$: contractive, $\alpha = 0,5$: conservative, $\alpha > 0,5$: divergent. Recommended: $0,5 \leq \alpha \leq 0,7$.

with the help of one variable of the dataset. The split which maximises the

reduction in impurity, based on squared probabilities of membership for each target category [64], is chosen according to a split target. This split target is called the Gini diversity index. With $RF(C_j, S)$ denoting the relative frequency of cases in the training set S that belong to class C_j, the Gini index is defined as:

$$G(S) = 1 - \sum_{j=1}^{x} RF(C_j, S)^2. \tag{3.25}$$

The algorithm partitions the data until every leaf contains cases of a single class. So the Decision Tree will classify all training elements. This effect is called overfitting and well described by [41, 63]. Overfitting leads to a loss predictive accuracy in most application, so that the tree is pruned to reduce some of its structure. CART employs the "1 SE rule", which chooses the smallest tree whose estimated error is within one standard deviation of the estimated error of the best tree.

3.4.3 Artificial Neural Networks

An Artificial Neural Network (ANN) is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing. It involves a network of highly complex processing elements (neurons), where the global behaviour is determined by the connections between these elements and their parameters. The topology of Artificial Neural Networks varies from simple perceptrons (two layer neural networks) to high structured ANNs. Fig. 3.6 gives an example of both structures. The original inspiration for the technique is grounded in biology and refers to the examination of the central nervous system and its neurons. A neural network can be described also as the synergetic use of a non-linear function and a learning-algorithm. ANNs qualify for domains lacking explicit systematical (expert) knowledge of the problem to be solved [17].

The outstanding feature of ANNs is their ability of learning, which in practice means that a (data) set of observations is used to train the network in order to solve a certain task in an optimal sense. Also ANNs are marked by the ability to generalise very well. The resulting neural network models can be trained by either supervised or unsupervised methods, depending on the nature of the task. The supervised learning methods use sets of instances in the form $\langle x, y \rangle$, where y represents the variable to predict and x is a vector of features thought to be relevant to determine y. The goal of supervised

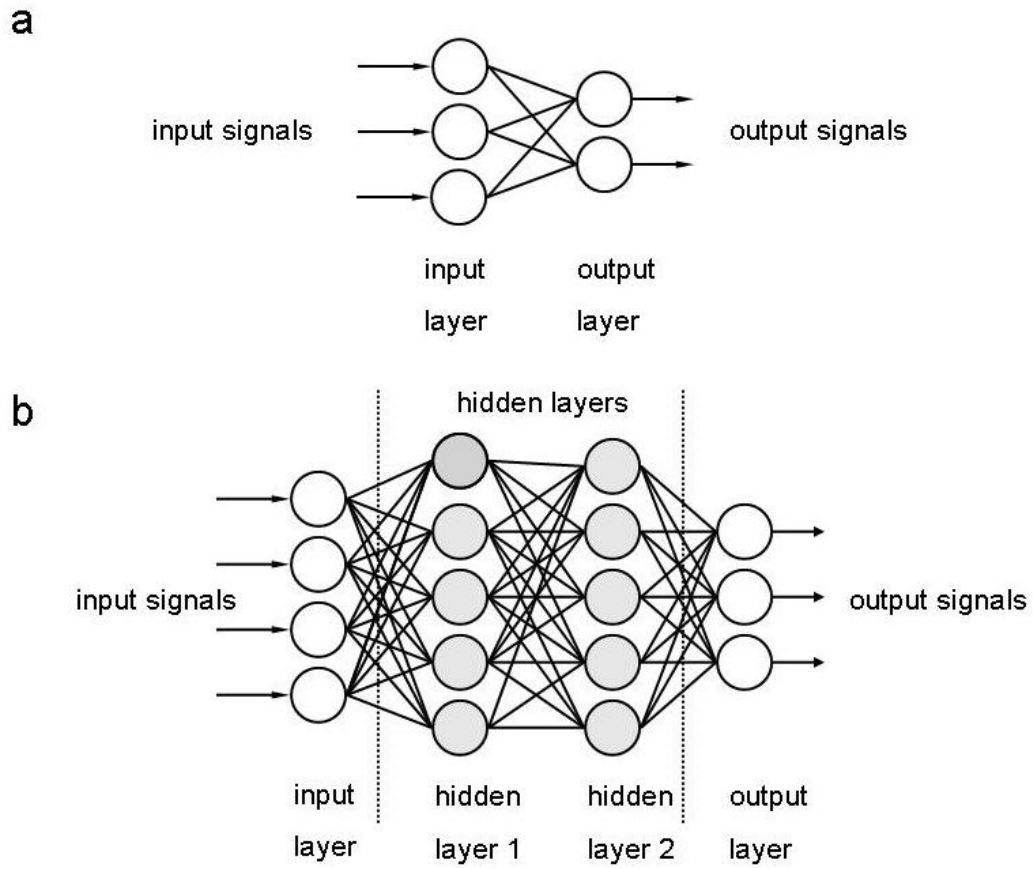


Figure 3.6: Structure of a simple perceptron (a) and a multi layer network with two hidden layers (b)

learning is to induce a model that allows to predict y values for previously unseen examples x. In unsupervised learning, the model is also build with the help of training examples, but each instance consists only of the x part without the y value. The goal of unsupervised learning is to build a model that accounts for regularities in the training set [28].

The papers [130, 2] give an overview on the Artificial Neural Networks and the learning methods, including the following:

- supervised learning
  - back propagation
  - superSAB

  - quickprop
  - resilient propagation

- unsupervised learning
  - hebbian learning

Artificial Neural Networks represent their learned solutions using real-valued parameters of the network (connection weights). Therefore, they are so called black box models, which means that their results can not be interpreted in a simple way. With respect to this, the main application of ANNs include control systems (vehicle control), pattern recognition (radar systems, face identification, object recognition), sequence recognition (gesture, speech, handwritten text) and financial applications. The application of ANN for Data Mining tasks like knowledge discovery or rule extraction is described by Craven [28].

The investigations carried out at the chair of Fluidmechanics and Process Automation of Professor Delgado show the aptidue of ANNs for prediction tasks of, for example, fermentation processes or the recognition of damaged beverage crates [124, 101, 123, 11, 13, 12]. The results of those investigations prove the aptitude of ANNs for Data Mining tasks like the prediction tasks of the presented thesis.

3.4.4 Fuzzy Logic

L. A. Zadeh initiated Fuzzy Logic in 1965 [125, 127, 126]. An overview to this topic, the development of Fuzzy Logic and further details are given by [7, 16, 19, 131, 32].

Whereas classical boolean logic expresses its statements in binary terms, Fuzzy Logic extends the boolean logic with the concept of partial truth. This multivalued logic representation of truth in form of degrees of truth allows intermediate values to be defined between conventional evaluations like true/false, yes/no, high/low, etc. and is therefore nearer to real-life problems and statements. Also, Fuzzy Logic can deal with imprecise statements and linguistic knowledge from experts [106, 107, 9]. Fuzzy Logic is implemented in several domains including household appliances (washing machines, refrigerators), automobile subsystems (ABS, cruise control), air conditioners and elevators.

Fuzzy Logic represents membership in vaguely defined sets, allowing values between 0 and 1 (including borders) and in its linguistic form inprecise statements like "slightly", "less", "very". These fuzzy sets, for example describing the input and output variables of a system, are an extension of classical set theory, defining elements in a binary way, either belonging to a set or not. A simple example of a fuzzy set is given in fig. 3.7, showing the membership $\mu(x)$.

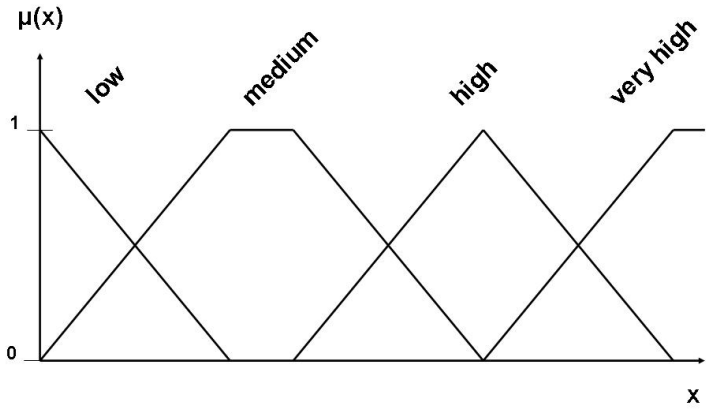


Figure 3.7: Simple Fuzzy Set

The rules are represented in the form

$$\text{if input is } A \text{ then output is } B. \tag{3.26}$$

Specifically, partial membership in a set is allowed. The relationship between fuzzy sets is expressed with "If-Then" rules [128, 121]. Building the fuzzy system and assessing the rules can be carried out in different ways [9, 42, 109, 67, 26].

Fuzzy sets own several features, approving this method for KDD purposes. Among these the user friendly linguistic terms excel, the rule-based system and concluding the descriptive and easy to interpret and visualise results.

At the investigators group of Professor Delgado, Fuzzy Logic is subject of many investigations. Several papers [5, 10, 13, 14, 57, 58, 69, 70] show the aptitude of those method for optimisation and prediction tasks.

In the presented work, the fuzzy modell is constructed using as well experts knowledge as results attained from other Data Mining algorithms. The resulting model is used for prediction of filtration, as similar approaches yielded good results [58, 57, 59, 84, 69, 5, 70].

3.4.5 EM Algorithm

The Expectation Maximisation (EM) algorithm is a broadly applicable and applied approach to the iterative calculation of maximum likelihood estimates. The algorithm consists of two steps at each iteration, called expectation step (E-step) and maximisation step (M-step). The name EM algorithm, deduced from these steps, was given by Dempster, Laird and Rubin in their fundamental paper [30]. A good overview on this algorithms, its features, limits and applications is given by [80, 115]. Among the applications of this algorithm are protein classification [78], modelling fish stocks [27] and modelling internet traffic [76]. The algorithm is also used for clustering purposes as described in [119].

The EM algorithm can be used to analyse superposed distributions. Within an observed single variable in a large sample, this algorithm identifies the underlying normal distributions of possible clusters within these observations. An example of two normal distributions with different means and deviation forming one superposition distribution is shown in fig. 3.8.

In this work, EM algorithm is applied to analyse data and to determine boundaries of groups with similar features. With a given distribution of observations, the algorithm is used to identify the means and deviations of n underlying distributions. In this approach, the E-step sets n distributions with estimated means and deviances. The M-step varies from the common used algorithm, as in this work the minimisation of the sum of squared errors is applied, in opposition to the maximisation of likelihood. Therefore, the EM algorithm approximates the observed distributions of valued based on mixtures of different distributions in different classes.

3.4.6 Principal Component Analysis (PCA)

In the presented thesis, PCA, a popular technique in pattern recognition, is applied for Knowledge Discovery from laboratory data. PCA is a technique that can be used to simplify a data set, by yielding a new coordinate system for the data setsuch that the greatest variance by any projection of the data

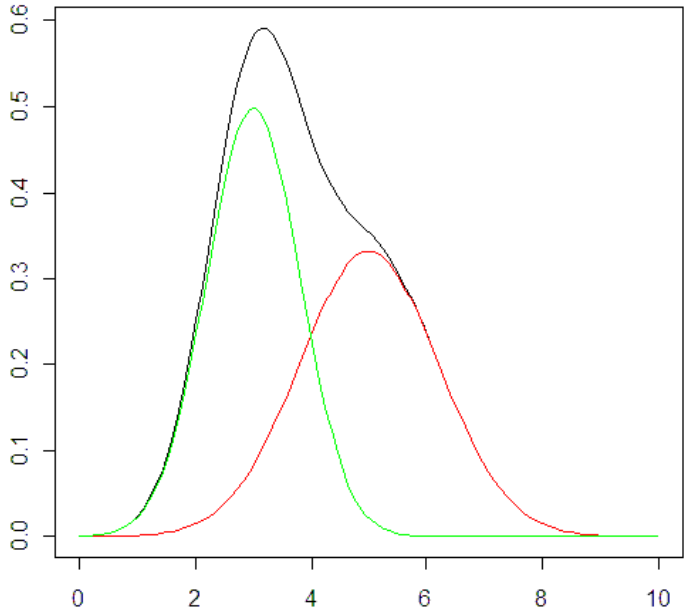


Figure 3.8: Example of superposition of two distributions

set comes to lie on the first axis, called the first principal component PC_1, the second greatest variance on the second axis PC_2 and so on. This is reached by a linear transformation of the data set. Thereby, PCA is used to reduce dimensionality in a data set while retaining those characteristics of the data set that contribute most to its variance. Thus, the low-order components contain the most important aspects of data.

The resulting principal components, subsequently denoted as loadings, represent eigenvectors of the data set matrix and are obtained by the determination of the eigenvectors and eigenvalues of the correlation matrix. Since the data matrix X can be constructed from a linear combination of the eigenvectors, a transformation between the matrix containing the loadings V and the original data matrix X must exist and is supposed to have the form

$$X_{n\times m} = S_{n\times m}\dot{V}_{n\times m}{}^{T} \tag{3.27}$$

where S is the score matrix containing the coefficients for the linear combination of the loadings.

The graphical representation of this methods supports the investigation of correlations and dependencies in data. The resulting diagrams project the initial variables into the subspace defined by the reduced number of load-

ings. Therefore, two principal components are plotted against each other. The length of the arrows representing the initial variables is proportional to standard deviation. The scalar product of two vectors is proportional to covariance of both variables. The cosines of the angle between to vectors represents the correlation coefficient for both variables. Further details on PCA are given by [46, 79, 51, 112]

3.5 Discovery System

To solve the defined task, software is needed to carry out Data Mining tasks. For this purpose several solutions are available, ranging from a set of specialised programs to a complete suite, including as well the needed Data Mining methods as data preprocessing and preparation. The different systems can be defined according to simple dimensions as summarised as follows.

- Supported tasks
  - single mining task, only one method
  - single mining task, multiple methods
  - multiple mining tasks
- Supported process steps
  - only Data Mining
  - also pre- and postprocessing
- Supported domains
  - generic system
  - domain specific
- Tool integration approach
  - macro integration
  - micro integration
- Architectural layers
  - data management
  - data aggregation

- pattern and model evaluation
- search
- user interaction

Various reports, including commercial ones, provide an extensive overview and evaluation of KDD tools, such as Two Crows, Aberdeen Group and Gartner Group. Whereas those reports run the risk of being out of date, some sources of information are regularly upgraded like KDD nuggets [53]. The most known Data Mining suites include IBM's Intelligent Miner, Oracle's Darwin, SAS's Enterprise Miner, Silicon Graphic's MineSet and SPSS's Clementine. Detailed information on these discovery systems is provided by [61, 104, 21, 100].

For the approach in the present work the open source software "R" [111] was chosen. R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It is based on the same programming language as S-Plus, a commercial statistic suite, and among other features providing a full reference of this language. R is widely used in academic environment and many packages, provided also as open source, extend the main package. Therefore, also packages can be adapted or extends to special needs and new functions can be implemented respectively. Additionally, for ANN and Fuzzy Logic an existing code, developed at the Chair of Fluidmechanics and Process Automation and the InformationTechnology Weihenstephan, is used.

The use of free software enables the compilation of a very well adapted and lean software suite for the application as a tool in breweries.

3.6 Evaluation

Within the iterative process of knowledge discovery the built model or models have to evaluated. A key objective is to determine whether the defined task and subtasks have been sufficiently considered. Therefore the model has to be tested on scenarios or real applications. Moreover, this evaluation process assesses not only the results necessarily related to the original objectives but also other Data Mining results generated, like findings which are not related to the objectives, but may also unveil additional challenges, information or hints for future directions. Recapitulating (see also [122, 6]), the results can be defined as

$$\text{results} = \text{models} + \text{findings}. \tag{3.28}$$

Both parts of this equation have to be evaluated. The main aspects of model evaluation are accuracy and generality. On the one hand, the models have to be accurate to fullfill the requirements, e.g. prediction tasks, but on the other hand they have to be general, because the models are built with past experiences but have to be applied to new datasets that will arise in the future. The classical technique for evaluation is to assemble a sample of cases that is devoted exclusively to evaluation. Fig. 3.9 shows the splitting of the data in two samples, one for training and one for evaluation.

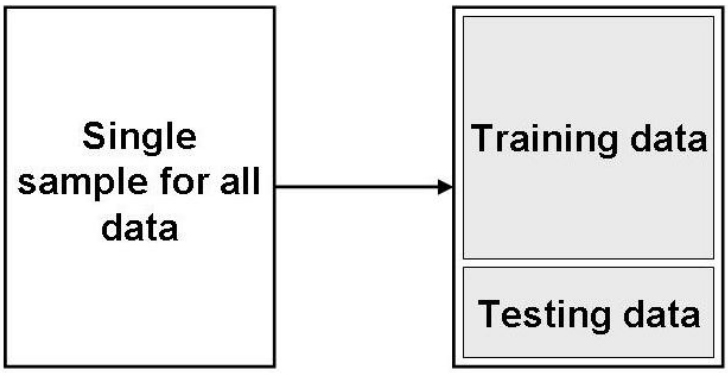


Figure 3.9: Data assembling for empirical evaluation

Relying only on training data would probably tend to exceptional results, being perfect for the used training data, but may not generalise to new examples. This overfitting or memorising of the model can be prevented by hiding the test datasets and comparing the training results to the results of the prior unseen datasets. In this work the training data covers 80% of the available data, whereas the remaining 20% is used for evaluation.

To determine the quality of the model with the given training and testing data, in this work the performance of prediction is measured. Unless a perfect model is found, the model will generate wrong predictions. Therefore the known labeld data, the testing data, is compared to the predictions and the error is measured. The error from classification is defined as

$$Error\,rate = \frac{number\,of\,errors}{number\,of\,examples} \quad . \tag{3.29}$$

A more detailed view of error can be reached by identifying the errors as false

positives (FP) and false negatives (FN). Based on a binary classification ($true$ / $false$), false negatives are those elements that were classified as $false$, but in fact are $true$, whereas the false positives are elements classified as $true$ but being $false$. FP and FN have an important influence on model quality, depending on the application. For instance in medical applications an erroneous diagnosis may have enormous consequences and therefore one may accept many more FPs then FNs for screening tests. FP and FN form together with the true positives (TP) the $precision$ (3.30) and $recall$ (3.31). Sometimes, they are combined to F, the harmonic mean of precision and recall (3.32) [117].

$$precision = \frac{TP}{TP + FP} \tag{3.30}$$

$$recall = \frac{TP}{TP + FN} \tag{3.31}$$

$$F = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \tag{3.32}$$

Secondly, the findings of the KDD process have to be evaluated. The question is, whether the results are novel and usefull. Depending on those aspects and on the sufficiency to fullfill the task, the process has to be reviewed. New objectives can arise from the findings or new subtasks can get important. These aspects have to be considered in the iterative KDD process.

Chapter 4

Results

Within this chapter the results of the KDD process, starting with the two classification approaches, followed by the analysis of laboratory data, the knowledge discovery and the approaches to prediction of filtration, are presented. Some explanation on methodology is included, provided that it was not introduced in chapter 3.

4.1 Classification of Filtration - First Approach

This approach is carried out to give a first insight into filtration. Because of the high reduction of data, several constraints emerge. Therefore, filtration having different states of the filter can not be compared. Solemnly this reduces the filtration to those ones which are operated on a clean filter. All following filtration of a batch start with an undefined filter. In the remainder of this section the data reduction, Cluster Analysis and the evaluation of the resulting clusters, followed by interpretation and comparison of the results are presented.

4.1.1 Reduction of Data

With respect to experts knowledge not all of the available data is used to characterise filtration, but only five attributes are concerned. These are:

- haze measured at 25° (h_{25}),
- haze measured at 90° (h_{90}),

- volume flow ($\dot{V}$),
- pressure difference, calculated from inlet and outlet pressure ($\Delta p = p_{out} - p_{in}$) and
- kieselguhr, given by its dosage (d_{KG}) and the overall amount, used for one filtration ($\sum KG$).

With the exception of $\sum KG$ the attributes are time series, with values measured normally every six seconds. Figure 4.1 shows these time series of a sample filtration. Whereas $\dot{V}$, d_{KG}, h_{25} and h_{90} show alternating curves, the value of Δp increases within the batch more or less constantly, with respect to the principles of filtration and the regulation of d_{KG} and $\dot{V}$.

The data reduction is carried out by narrowing down the time series to just one value. Therefore, the data of $\dot{V}$, h_{25} and h_{90} is reduced to the respective arithmetic mean $\bar{\dot{V}}$, $\bar{h}_{25}$ and $\bar{h}_{90}$. These means are also shown in Figure 4.1, diagram a, d and e. The attribute kieselguhr is considered by overall applied amount of the filtration $\sum KG$. The pressure difference is expressed by the overall difference $\Delta p^* = \Delta p_{end} - \Delta p_{begin}$.

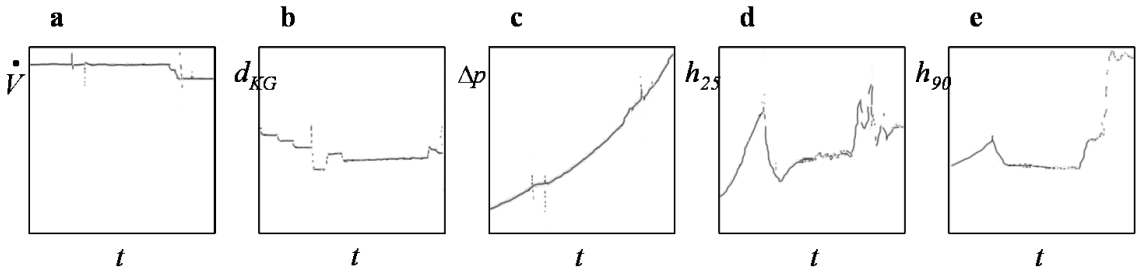


Figure 4.1: Diagrams of a) $\dot{V}$, b) d_{KG}, c) Δp, d) h_{25} and e) h_{90}

4.1.2 Cluster Analysis

Based on the reduced data described in the section above, four different Cluster Analysis methods are applied. According to Chapter 3.4.1 these are the two hierarchical methods *AGNES* and *DIANA* and another partitioning one, *PAM*. More over, the parameters of the cluster methods are evaluated to estimate the most applicable settings, e.g. which agglomerative criterion yields the best result.

The different algorithms are firstly applied to the data of filtration of light lager beer with a cleaned filter, because as described in chapter 3.1.1 a comparison with filtration processed on an used filter is not possible.

AGNES

The Agglomerative Nesting Cluster Method is influenced by three parameters. These are the metric used for calculating dissimilarities, the clustering method and the standardisation of values. As all data is standardised with the data preprocessing, this option is not used. The distance of elements is always calculated with the euclidian metric. Therefore, only the clustering method is varied. The agglomerative coefficient *ac* is used to compare the quality and applicability of the results. Fig. 4.2 compares the *ac* of the different parameters.

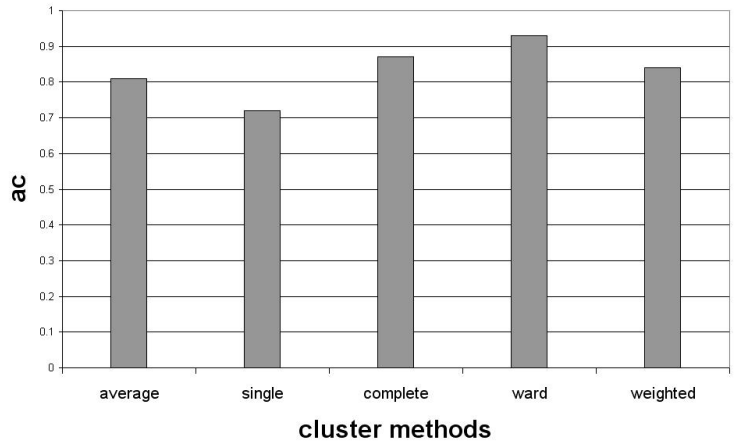


Figure 4.2: Comparison of AGNES cluster methods

The method yielding the best coefficient is Ward's method. The dendograms of the different method are displayed in Appendix B.1. They demonstrate some of the characteristics described in Table 3.6, especially the chaining effect of the single linkage method can be seen in Figure B.2.

DIANA

This divisive cluster algorithm results a divisive coefficient $dc = 0.87$. The *dc* is the counterpart of the *ac* and can be compared to it [56]. Variation is not necessary, as only one method is available. Furthermore, the Euclidean

distance and no further standardisation are applied. Figure B.6 presents the dendogram of the clustering.

PAM

Apart from standardisation and metric, which are handled as described above, this partitioning method needs only the number of clusters as parameter. In opposition to the hierarchical methods the results of this method are not displayed as a dendogram but with a silhouette plot and characterised by the average silhouette width asw and the silhouette width of the clusters s_i. The plots for $n = 3 \ldots 9$ clusters are shown in Figure B.7 – B.13.

The comparison of the results of the PAM algorithm are displayed in Figure 4.3. The heighest asw and therefore the best results yield the approaches with $n = 3$ and $n = 4$, but the according silhouette plots (Fig. B.7 and Fig. B.8) indicate at least one very inhomogeneous cluster within the clusterings of those approaches.

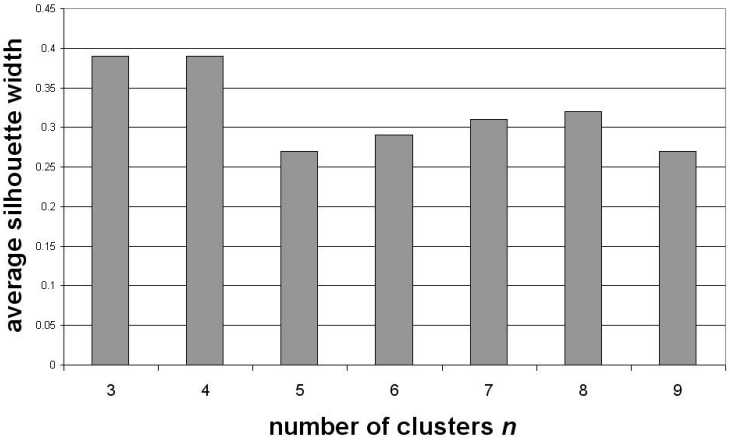


Figure 4.3: Comparison of asw for $n = 3 \ldots 9$ Clusters

4.1.3 Conclusion

Recapitulating, Ward's method of the agglomerative approach yields the best result. The other methods of AGNES succumb to Ward's method and also the result of the divisive approach. The partitioning algorithm divides the dataset with respect to a given number n of clusters, but the clusterings of these approaches are very inhomogeneous.

To verify the thesis of Ward's method being the appropriate one for this task, it is applied to further beer types. Figure 4.4 shows the resulting *ac*'s of clustering the data from four beer types, two different light lager, a Pilsener and a nonalcoholic beer. These are the beer types commonly processed on a clean filter. Very homogenous clusters result from this approach, having *ac*'s from 0.94 – 0.96. The dendograms are displayed in the Appendix (Fig. B.14 – B.17).

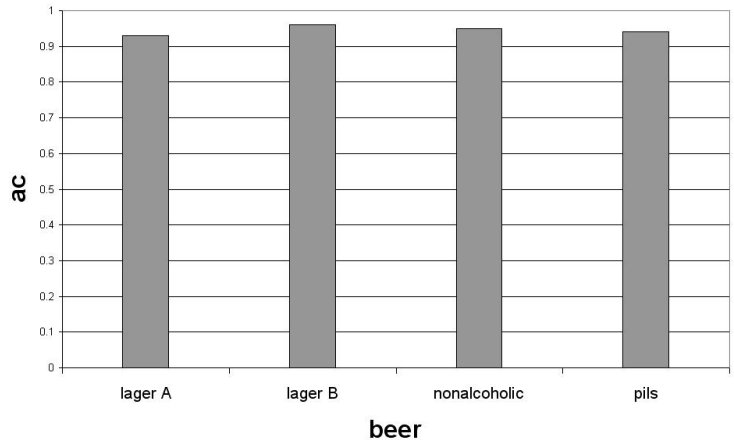


Figure 4.4: Ward's method applied on Filtration Data of 4 different beer types

A remaining question with Ward's method is defining the number of clusters, as the hierarchical methods split the dataset until each the number of clusters is equal to the number of elements with each cluster having one element. Therefore, the number of clusters depends on the desired homogeneity. According to the height, respectively the level of distance as described in Chapter 3.4.1, a level has to be chosen to define the clusters. The lower they are defined, the more clusters, being very homogenous, are generated. Therefore, a compromise about accuracy and number of clusters has to be effected. Depending on the beer type cutting at a level of about 0.2 – 0.4 produces approximately seven clusters. With respect to the respectively dendogram the clusters are generated by cutting at this level, aiming at 5 to 9 clusters. Resulting groups of at least 3 elements are aspired, as well as groups showing equal numbers of elements. Thus, for the example of the light lager beer the level to separate clusters is chosen at 0.25, yielding 8 clusters form which only one contains a single element.

4.1.4 Evaluation of Clusters

The clustering itself produces unlabelled groups. These groups include similar elements differing from those of other groups, with the heights of the dendogram measuring the difference. In this first approach the labelling of the groups is carried out with respect to the elements attributes. As mentioned in Chapter 4.1.1, the used filtration attributes are the arithmetic means of volume flow and haze $\bar{\dot{V}}$, $\bar{h}_{25}$ and $\bar{h}_{90}$ as well the used amount of kieselguhr $\sum KG$ and the overall pressure difference Δp^*.

To compare the groups the arithmetic mean of each attribute for the complete dataset and furthermore for each cluster is computed (Table 4.1). A complete overview on the clustering with respect to each attribute is given in Appendix B.18 - B.22.

Table 4.1: Overall Means and Cluster Means of *Lager A*

	n	$\mu_{\bar{h}_{25}}$	$\mu_{\bar{h}_{90}}$	$\mu_{\bar{\dot{V}}}$	$\mu_{\Delta p^*}$	$\mu_{\sum KG}$
overall	56	0.304	0.226	0.585	0.067	0.174
1. cluster	3	0.352	0.528	0.576	0.063	0.222
2. cluster	10	0.375	0.256	0.579	0.059	0.175
3. cluster	6	0.260	0.293	0.579	0.075	0.161
4. cluster	3	0.515	0.321	0.567	0.070	0.212
5. cluster	6	0.496	0.173	0.562	0.084	0.172
6. cluster	21	0.245	0.169	0.593	0.056	0.161
7. cluster	6	0.150	0.181	0.603	0.097	0.155
8. cluster	1	0.086	0.143	0.602	0.064	0.390

The cluster means are compared to the overall mean of all elements. A binary rating system results by labelling each cluster with "0" for a cluster mean *better* than the overall mean and 1 respectively. The result *better* is defined by technological view as means less than the overall mean for h_{25}, h_{90}, Δp^*, and $\sum KG$ and a higher cluster mean for $\dot{V}$. Thus, each cluster is described by the binary labels $LB_{\bar{h}_{25}}$, $LB_{\bar{h}_{90}}$, $LB_{\bar{\dot{V}}}$, $LB_{\Delta p^*}$, and $LB_{\sum KG}$ as shown in Table 4.2.

In the following, the binary labels are combined to a five digit binary number with the first, left digit representing $LB_{\bar{h}_{25}}$ up to the last digit covering $LB_{\sum KG}$, according to the columns of Table 4.2. A further simplification is the transfer from the binary labels to a rating based on one decimal number. Therefore, the 5 binary labels of one cluster are merged to one number by

Table 4.2: Binary Labels of Clusters of *Lager A*

	n	$LB_{\bar{h}_{25}}$	$LB_{\bar{h}_{90}}$	$LB_{\bar{V}}$	$LB_{\Delta p^*}$	$LB_{\sum KG}$
overall	56	-	-	-	-	-
1. cluster	3	1	1	1	0	1
2. cluster	10	1	1	1	0	1
3. cluster	6	0	1	1	1	0
4. cluster	3	1	1	1	1	1
5. cluster	6	1	0	1	1	0
6. cluster	21	0	0	0	0	0
7. cluster	6	0	0	0	1	0
8. cluster	1	0	0	0	0	1

adding the values in a decimal way. Thus, a range from $0 \dots 5$ results, with the rating showing the number of attributes being poorer than the overall mean of filtration of the corresponding beer.

Adjacent, the classes are defined as better filtration (rating $0, 1, 2$) or as poorer filtration (rating $3, 4, 5$) compared to the mean of all filtration for one beer type. Table 4.3 shows the evaluations for the example of one beer type, the evaluations of the remaining beer types are given in Tables B.1 - B.3. As Figure 4.5 shows, the number of poorer and better filtration is the same.

Table 4.3: Evaluation of Clusters of *Lager A*

	binary label	**rating**	**description**
1. cluster	11101	4	poorer
2. cluster	11101	4	poorer
3. cluster	01110	3	poorer
4. cluster	11111	5	poorer
5. cluster	10110	3	poorer
6. cluster	00000	0	better
7. cluster	00010	1	better
8. cluster	00001	1	better

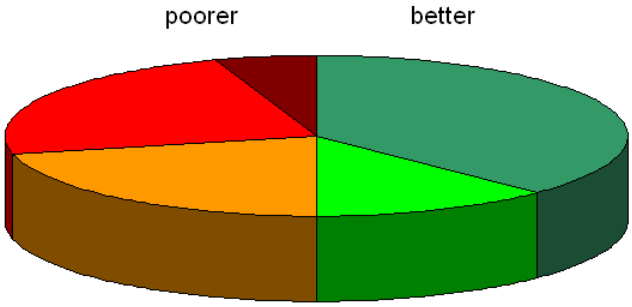


Figure 4.5: Evaluation of Clusters of *Lager A*

4.1.5 Interpretation of Clusters with Decision Tree

The cluster algorithm carried out above yields similar groups but gives no insight into the underlying reasons or aspects for the clustering. In this approach a Decision Tree algorithm is used to reveal structures of clusters.

Thus, the datasets already used for clustering are expanded by the evaluation defined in section 4.1.4. The Decision Tree method divides the data as described in section 3.4.2 and yields the attributes and values used for splitting. In this case it is the objective of the method to split the datasets according to the evaluation. The results are shown in B.23 - B.26, the rating from $0 \ldots 5$ is used for analysis.

The decisions trees show a major influence of haze values, both h_{25} and h_{90}, on the clusterings. Minor impact is applied by kieselguhr and pressure difference.

4.1.6 Comparison of the Results with Experts Knowledge

To verify the method and its applicability the results are compared with knowledge of experts. The reference brewery positions thresholds for each of the five attributes and for each type of beer, describing different qualities. The threshold values are anonymised, as well as the attribute values of filtration are already standardised (see Chapter 3).

Table 4.4 shows the comparison of threshold values for the h_{25} attribute of bottom-fermented beer (expert knowledge) with limits found by the Data Mining methods.

Table 4.4: Comparison of h_{25} thresholds from experts knowledge and Data Mining results

	experts knowledge	**Data Mining result**
very good	$< x_1$	–
good	$< x_2$	$< 0.9 \cdot x_2$
middle	$x_3 - x_4$	$> 1.04 \cdot x_3$
poor	$> x_4$	$> 1.13 \cdot x_4$

4.2 Classification of Filtration - Second Approach

This second approach of classifying filtration aims at improving the classification compared to the first one. Especially a better delineation of filtration is aspired. Moreover, the restriction to only the first filtration, having a defined status, is avoided, but irrespective of preceding batches filtration can be taken in account. Therefore, a more complex way of describing filtration is applied as explained in the following. This approach considers the time dependent behaviour. Furthermore, dummy filtration are introduced, helping to evaluate filtration.

4.2.1 Reduction of Data

Basically, the same attributes are used as in the first approach: h_{25}, h_{90}, $\dot{V}$, Δp^* and (d_{KG} or $\sum KG$, respectively. In contrast to the first approach data is reduced not to arithmetic means but to monotone functions. This is carried out by integration or summation of process data or by the combination of them according to technological principles.

Moreover, in this approach the attributes are not related to (filtration) time but are transformed in relation to accumulated kieselguhr $\sum KG$. Expected advance of this approach is a better comparability of filtration due to the technological and technical background. Independently of volumetric flow and filtration time, the maximum amount of kieselguhr is given by the construction of the vessel. The remaining constraint, the pressure is dependent of the kieselguhr layer [35] and therefore of the amount of kieselguhr, too.

The data reduction of the two haze attributes is carried out with summation

of values. The following four figures explain the process on the example of h_{25} of one filtration. Figure 4.6 displays the raw data of this precise filtration. The typical characteristics of raw data are visible: noisy and missing data.

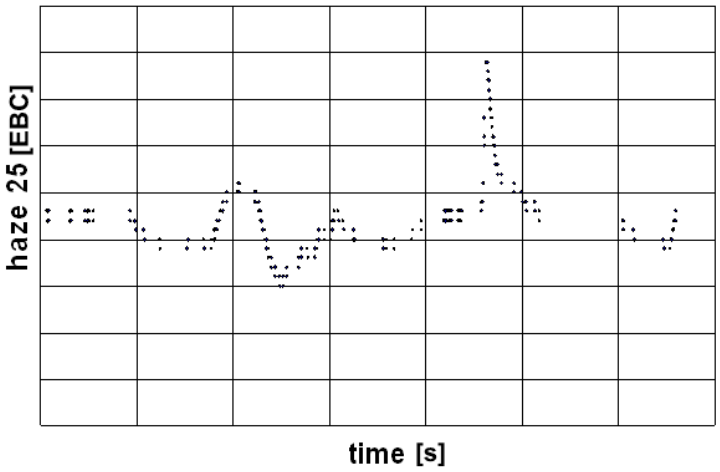


Figure 4.6: Example filtration: raw data of h_{25}

These effects are abolished by interpolation which is carried out with three objectives: removal of noise, replacement of missing data and synchronisation of the different attribute series. Further process steps need these synchronised data. The different sensors may have different time stamps and only after synchronisation a precise assignment and charging against each other is possible. The gaps of data, origin from sensors being out of synch or from missing data, are filled by a linear interpolation between the last and the next known sensor value. Thus, the unknown sensor value v_2 at the time t_2 is interpolated between the previous value v_1 at t_1 and the next value v_3 at t_3 as follows ($t_1 < t_2 < t_3$):

$$v_2 = \frac{v_3 - v_1}{t_3 - t_1} \cdot (t_2 - t_1) + v_1 \tag{4.1}$$

Fig. 4.7 shows the interpolated series of h_{25} of the example filtration. Noise and gaps are eliminated but still the alternating characteristic of the series complicates the classification. Therefore, in this approach the series of h_{25} and h_{90} are transformed by summation. It is emanated that the haze value of the suspension beer is correlated with the mass of the particles causing the haze. The summation is based on this assumption. Furthermore, for reasons of comparability the attributes are now related to the amount of kieselguhr.

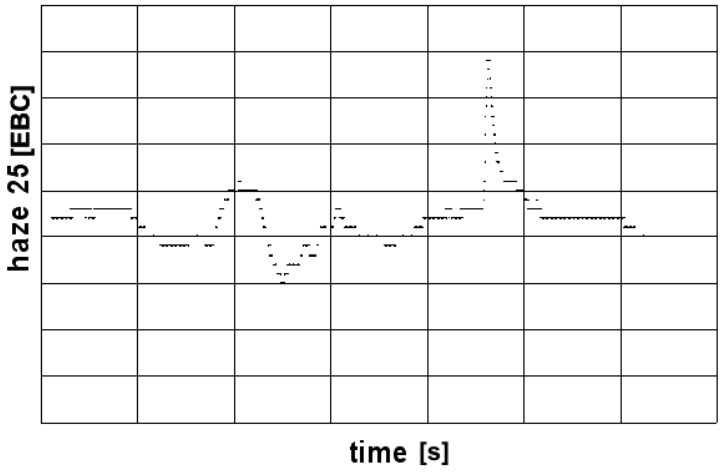


Figure 4.7: Example filtration: interpolated data of h_{25}

The third figure (Fig. 4.8) shows the summation of h_{25} dependent of the summation of kieselguhr. This new illustration yields a monotone increasing characteristic of the transformed attribute, being obviously nearly linear. Generality of the example filtration is shown in Appendix C (Fig. C.1). The characteristics of the series resemble those of the example.

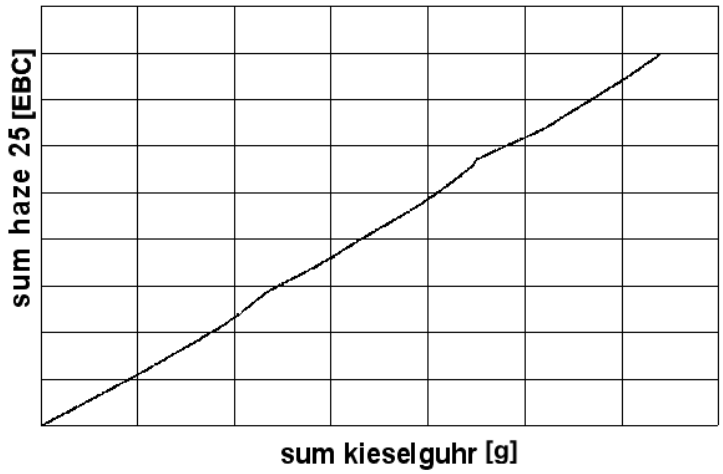


Figure 4.8: Example filtration: summation of h_{25}

The last diagram (Fig. 4.9) displays the approximation with a linear equation of the form $y = bx + c$ with b representing the slope and c the axis intercept

which is in this case the origin ($c = 0$), based on the technological fact that at the beginning of filtration ($\sum KG = 0$) no haze particles were applied on the filter. Already the summation shows this feature.

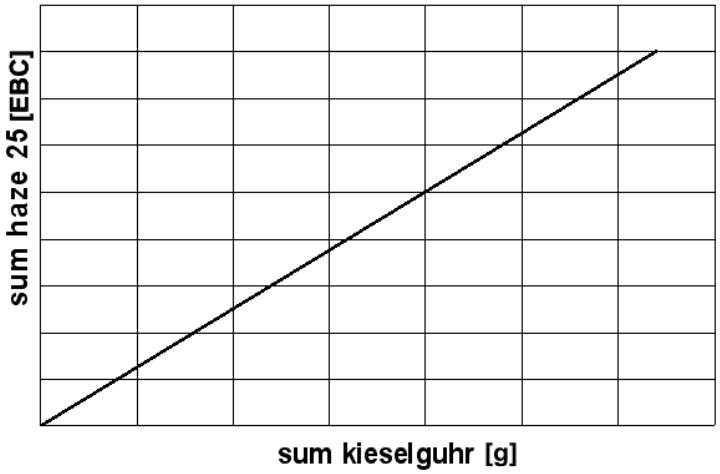


Figure 4.9: Example filtration: linear regression of h_{25}

In the Appendix C (Fig. C.1 and C.2) the summation and regression diagrams of several filtration are displayed. The representation of the time series by the transformed attributes and their linear equations is evaluated by the suitability of this approximation. An overview on the quality of the linear regression gives Fig. 4.10 which shows the mean of correlation and its deviance as well as minimum and maximum values for the regression of all filtration for the four beer types *Hell A* (176 filtration), *Hell B* (288 filtration), *Nonalcoholic* (82 filtration) and *Pilsener* (98 filtration). It can be seen, that most filtration can be approximated by linear regression with a very high correlation. Obviously, linear regression of summation qualifies very good for approximation of the haze time series.

The two remaining attributes $\dot{V}$ and Δp^* are treated in a different way due to different initial situation. On one hand, volumetric flow is kept constant during most of filtration time at the desired value and only regulated to adapt the process to filtration problems for example. On the other hand, pressure difference shows in most cases a monotonic characteristic, cause by the increasing kieselguhr layer and the residues within the layer. Therefore, summation is not taken into account for these attributes, but more over they are regarded with respect to technological facts and known dependencies which are introduced briefly in the following. Evers [35] gives a detailed introduction to this topic.

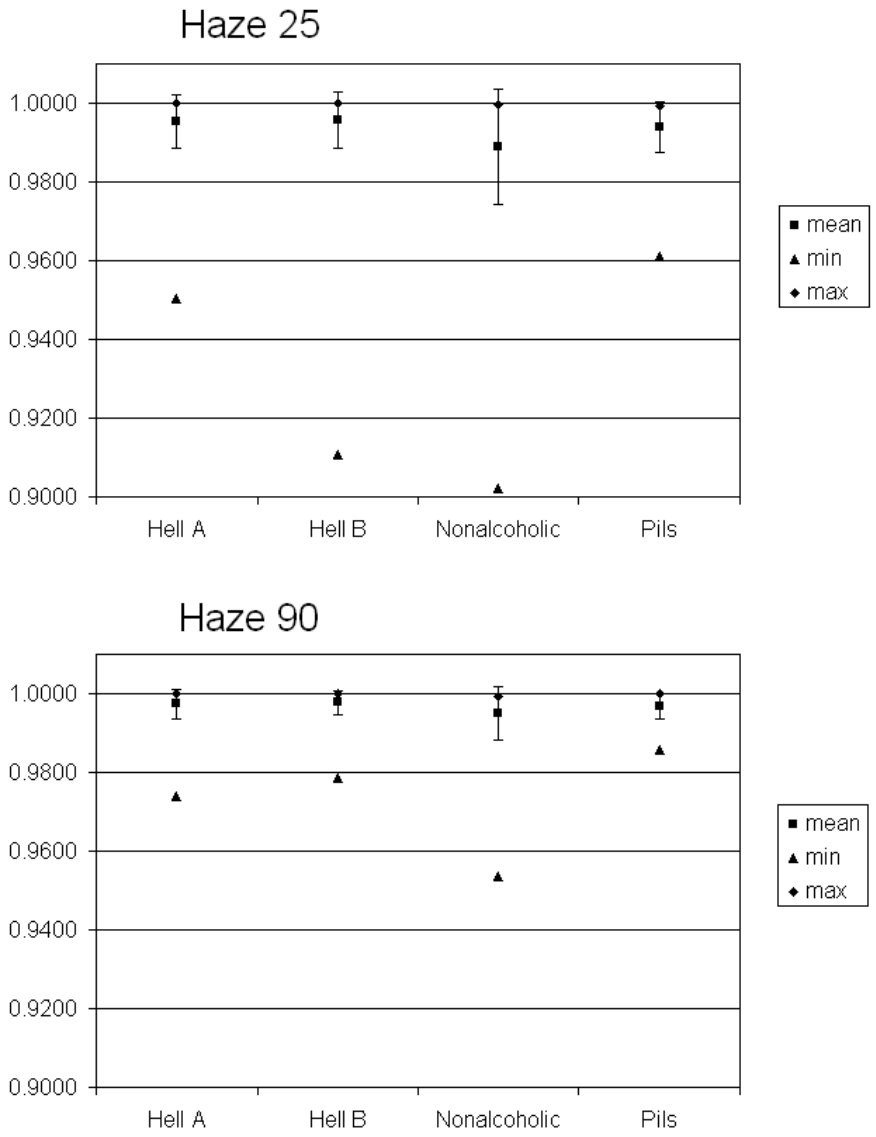


Figure 4.10: Comparison of correlation coefficient of linear regression of h_{25} and h_{90} for different beer types.

The ideal cake filtration is described by Eq. 4.2, known as Darcy's law. Stringently, this equation is suitable only for laminar flows of pure newtonian, non polar fluids through porous, incompressible filter cakes.

$$\frac{dV}{dt} = \frac{\Delta p \cdot A}{\eta \cdot R} \tag{4.2}$$

Thus, the volume flow is proportional to pressure difference between inlet and outlet of filter Δp as well as to the filter area A and inversely proportional to dynamic viscosity of the suspension η and filter resistance R. The latter parameter is made up of cake resistance r_c and resistance of the supporting layer r_0 as shown in Eq. 4.3 [35].

$$R = r_c + r_0 \tag{4.3}$$

Evers [35] describes also that the cake resistance increases proportionally to the layer thickness h_c (Eq. 4.4) inducing the specific cake resistance α.

$$r_c = \alpha \cdot h_c \tag{4.4}$$

With the assumption of A and η being constant Eq. 4.2 can be transformed as follows:

$$R \cong \frac{\Delta p}{\dot{V}}. \tag{4.5}$$

Moreover, from Eq. 4.3 and 4.4 it can be deduced, that the filter resistance R behaves proportional to the increasing filter cake and therefore to the amount of Kieselguhr:

$$R \cong \sum KG \tag{4.6}$$

Furthermore, Hagen-Poisseuille's law (Eq. 4.7) and a variation of the filter equation from Carmann-Kozeny (Eq. 4.8) can be applied for a better description of filter processes in brewery.

$$\frac{dV}{dt} = \frac{A \cdot \Delta p \cdot \varepsilon \cdot d_0^2}{\eta \cdot 32 \cdot h_c} \tag{4.7}$$

$$\frac{dV}{dt} = \frac{A \cdot \Delta p \cdot \varepsilon^3}{\eta \cdot h_c \cdot K' \cdot O_s^2 \cdot (1-\varepsilon)^2} \tag{4.8}$$

with

A filter area

Δp pressure difference

ε porosity

d_0 diameter of capillars

η viscosity

h_c layer thickness

K' Kozeny resistance constant

O_s specific surface

These coherences rely on generalising assumptions like the flow through parallel cylindrical capillaries (4.7) or on parameters like specific surface of filter cake O_s which can be measured only with much effort.

Therefore, in practice an enhanced Darcy equation (Eq. 4.9) is used. This equation considers cake resistance as well as resistance of supporting layer and consists of measurable and derivable quantities, respectively [35].

$$\Delta p(t) = \frac{\eta}{A} \cdot \left(\frac{\alpha \cdot \kappa}{A} \cdot V(t) + \beta \right) \cdot \frac{dV}{dt} \tag{4.9}$$

In the following the transformation of Δp and $\dot{V}$ is carried out analogous and according to Darcy's law and the equations 4.5 and 4.6.

Consecutive, $\frac{\Delta p}{\dot{V}}$ is illustrated dependent of $\sum KG$ as figure 4.11 shows.

In agreement with the equations mentioned above, this figure describes the increment of filter resistance, for this example approximated with a linear regression. In this case the intercept denotes the filter resistance at the beginning of filtration. The slope illustrates the increasing resistance. In this special case the linear regression yields a good approximation of the characteristic but in the following also nonlinear regression considering a squared term $y = ax^2 + bx + c$ is applied. Generally, the axis intercept defines a measurement that enables the conjointly analysis of all filtration, regardless of the number of previous filtration.

Just as an overview, Figure 4.12 shows filtration of the pale ale *hell A* carried out on a freshly cleaned filter. With only few outliers the filtration feature low resistances at the beginning, evolving further on with differing characteristics

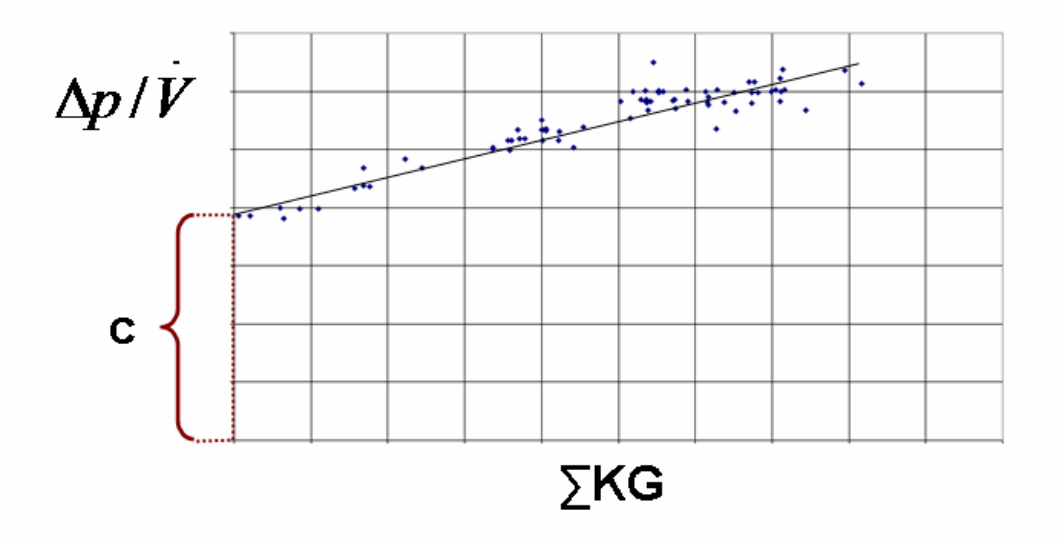


Figure 4.11: Graphical representation of Δp and $\dot{V}$ dependent of $\sum KG$ (example filtration)

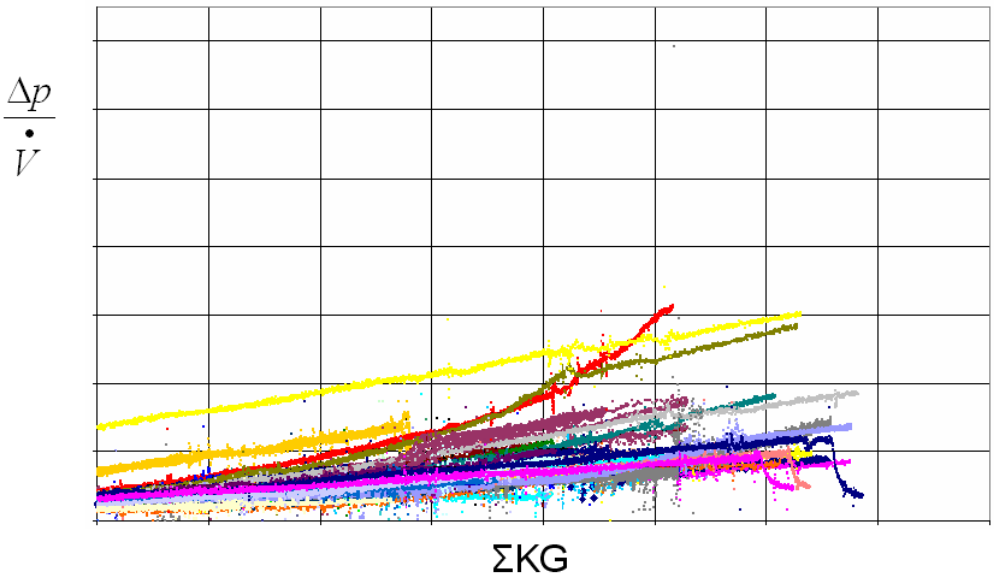


Figure 4.12: Δp, $\dot{V}$ dependent of $\sum KG$ regarding first filtration of *hell A*

pertaining length of filtration and increase of the characteristics. Figure 4.13, using the same axis scaling, displays filtration of the same beer applied on the filter as second batches. Dependend on antecedent filtration and their

impact on the filter, the second filtration show different initial conditions with a wider range than the first ones. Also, on this two figures a problem of assignment of the batches becomes obvious, as some filtration show a strong decreasing characteristic at the very end of the process. The filtration showing this effect are found to be the last filtration of a process without any following ones. To avoid this effect in the following the filtration data, mainly the duration of filtration, is adopted to this fact.

As already mentioned above, for the approximation of this filtration linear ($y = bx + c$) and quadratic ($y = ax^2 + bx + c$) regression are applied. Figure 4.14 displays the results of the linear regression whereas figure 4.15 shows the ones of nonlinear regression. A deeper look at the correlation of both regressions reveals that the nonlinear regression does not enhance the linear one. This is underlined by the quadratic term a being negligible low compared to the linear factor b. Appendix C contains the comparison of correlation (Tab. **??**) and analysis of the quadratic term carried out with a distribution analysis (Fig. **??**). Because of the marginal enhancement reached by nonlineare regression, in the following the more simple approach of a linear regression is applied.

Thus, in the remainder the axis intercept c and the linear term b of regression are considered, with c describing the initial state of filter and b yielding the increase of filter resistance.

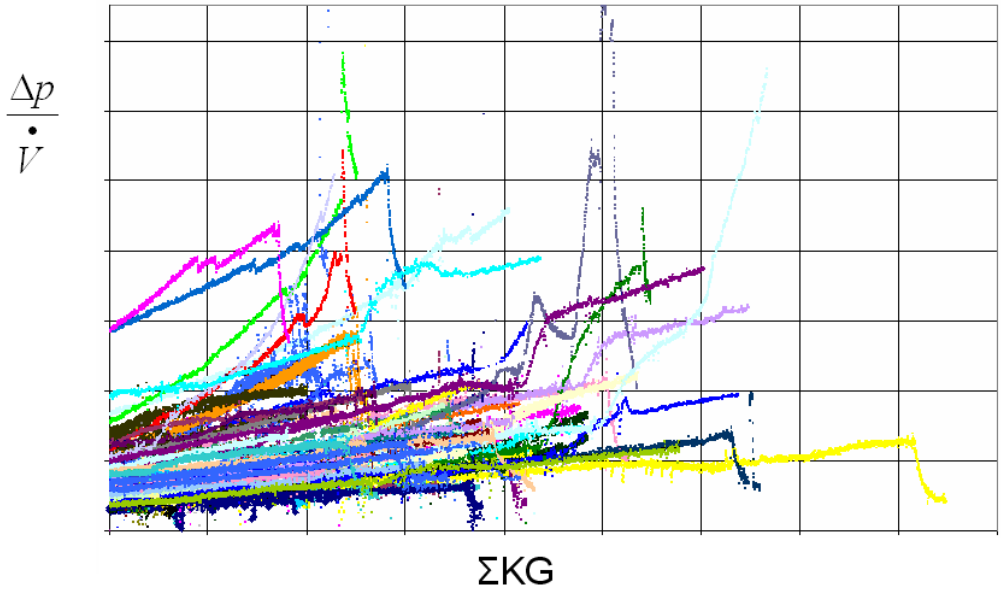


Figure 4.13: Δp, $\dot{V}$ dependent of $\sum KG$ regarding second filtration of *hell A*

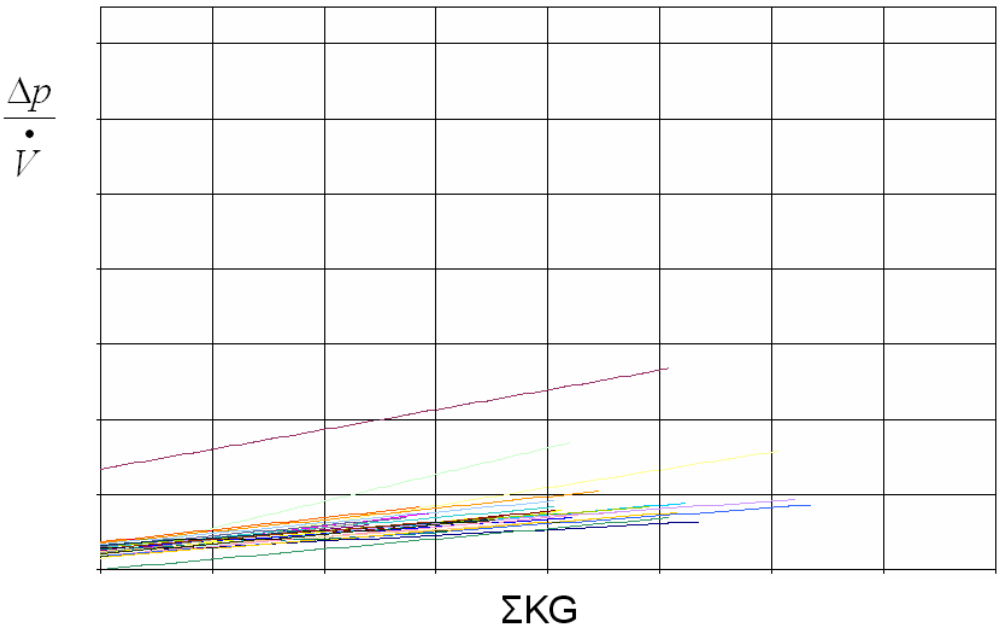


Figure 4.14: Linear regression of Δp and $\dot{V}$ dependent of $\sum KG$ regarding first filtration of *hell A*

4.2.2 Dummy Filtration

Furthermore, within this second approach dummy filtration are introduced. Dummy filtration represent virtual filtration characterised by the use of the threshold values which are considered by experts for evaluation of the process. Regarding these attributes and their thresholds, process data of filtration is generated by a Java program and appended to the respective database tables. The further processing like preprocessing or summation of dummy filtration is exactly the same as those of regular filtration data.

In detail, the attributes h_{25}, h_{90}, Δp and d_{KG} are considered with their thresholds shown by table 4.5.

These thresholds are upper limits so that a filtration is evaluated good regarding h_{25} with $h_{25} \leq h_{25}^{g}$. For a better representation of the evaluation ranges in this approach the values describing the mean of these ranges are applied, as described in table 4.6. As the *good* range has no lower limit, the therotical lowest value is considered, in case of the example $h_{25}^{0} = 0$. The other attributes are treated analog.

The remaining attributes needed for dummy generation are volume flow $\dot{V}$

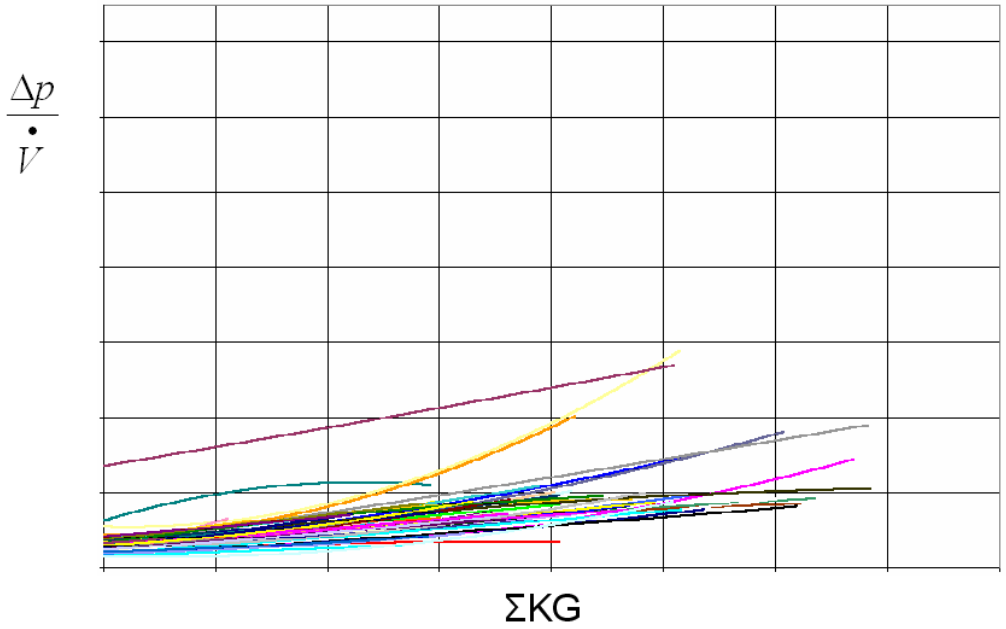


Figure 4.15: Nonlinear regression of Δp and $\dot{V}$ dependent of $\sum KG$ regarding first filtration of *hell A*

Table 4.5: Attributes and their thresholds from experts knowledge

evaluation	h_{25}	h_{90}	Δp	d_{KG}
good	h_{25}^g	h_{90}^g	Δp^g	d_{KG}^g
middle	h_{25}^m	h_{90}^m	Δp^m	d_{KG}^m
poor	h_{25}^p	h_{90}^p	Δp^p	d_{KG}^p

Table 4.6: Means of evaluation ranges as used for generation of dummy filtration with the example of h_{25}

evaluation	**lower limit**	**upper limit**	**mean of range**
good	h_{25}^0	h_{25}^g	$h_{25}^{g*} = 0.5 * (h_{25}^g - h_{25}^0) + h_{25}^0$
middle	h_{25}^g	h_{25}^m	$h_{25}^{m*} = 0.5 * (h_{25}^m - h_{25}^g) + h_{25}^g$
poor	h_{25}^m	h_{25}^p	$h_{25}^{p*} = 0.5 * (h_{25}^p - h_{25}^m) + h_{25}^m$

and output pressure p_{out}. As the maximal volume flow $\dot{V}^{max}$ is aspired, this value is used to compute the process data. Furthermore, of all filtration

output pressure is found to be p^c_{out} with only very little deviation so this value is also regarded for dummy filtration.

As the flow of kieselguhr suspension $\dot{V}_{KG}$ $[l/h]$ is considered by filtration PDA it is calculated as follows (Eq. 4.10):

$$\dot{V}_{KG}[l/h] = \frac{d_{KG}[g/hl] \cdot \dot{V}^{max}[hl/h]}{c_{KG}[kg/l]}. \tag{4.10}$$

For all these attributes process data is computed considering a filtration of one hour. These process data are stored analog to the regular process data and also processed similar, afterwards. As within this approach the linear term of regression of those transformed attributes is taken into account, the filtration time of one hour suffices.

In the following these induced, virtual filtration based on experts knowledge are used as reference processes. They act as tracers and facilitate the evaluation of the adjectant methods and their results.

To visualise and compare dummy and regular filtration a web diagram is considered. Figure 4.16 show the dummy filtration marked with the colours green, yellow and red as they are displayed in general.

This diagram displays the attributes of dummy and regular filtration in four dimensions. These are the slopes of regression of haze values $b_{h_{25}}$ (slope of $\sum h_{25}$ dependent of $\sum KG$) and $b_{h_{90}}$ (slope of $\sum h_{90}$ dependent of $\sum KG$) and the slope of increase of filter resistance $b_{\Delta p/\dot{V}}$ (slope of $\Delta p/\dot{V}$ dependent of $\sum KG$). The fourth dimension represents the ratio of applied kieselguhr $\sum KG$ and the filtrated volume of the corresponding batch V_{batch}. These four dimensions represent two features: quality and performace of filtration. The quality of the filtration is measured by the two attributes containing the haze values, as it is the target of filtration to produced a bright beer. The impact of the process on the filter as measured by filter resistance and the mass of applied kieselguhr diplays the performance of filtration.

Figure 4.17 shows some example filtration compared to dummy filtration using the web diagram.

These four filtration are representative ones of the whole universe of regular filtration. Filtration **d** shows very good quality and performance as all four attributes are lower than the ones of the good dummy filtration. Filtration **c** describes a filtration with a very low increase of filter resistance combined with medium ($b_{h_{90}}$) and high ($b_{h_{25}}$) slopes. For this example the rate of $\sum KG$ to V_{batch} attracts attention as it is very high. A deeper look at this

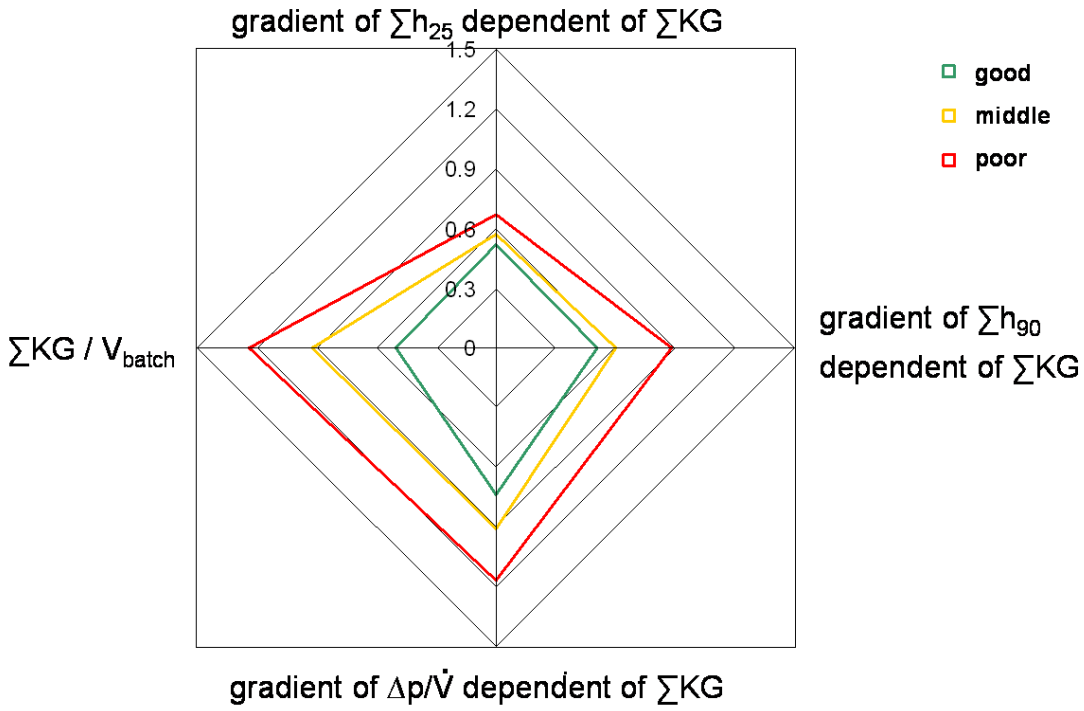


Figure 4.16: Visualisation of filtration using web diagram

filtration and its original characteristics explains this diagram. This precise filtration **c** is characterised by two features: the batch size of this filtration **c** is very small and it is the first batch on a clean filter. As the first batches are processed with declining dosage of kieselguhr emerging from a high dosage, this outstanding small batch shows a very high rate of kieselguhr compared to the filtrated volume. Whereas filtration **b** shows very high slopes and a medium rate of kieselguhr and volume characterising a very poor batch regarding both quality and performance, filtration **a** owns poor quality features combined with good to middle performance.

4.2.3 Cluster Analysis

Similarly to the first approach reduced data is used for further analysis, starting with Cluster Analysis. Following the results of the first approach, only *AGNES* is applied as hierarchical method, supplemented by the *k-means* algorithm as a partitioning method. Combined with representative dummy filtration the *k-means* algorithm is preferred over the earlier used method *PAM*, because of its better aptitude, explained by the use of dummy filtration

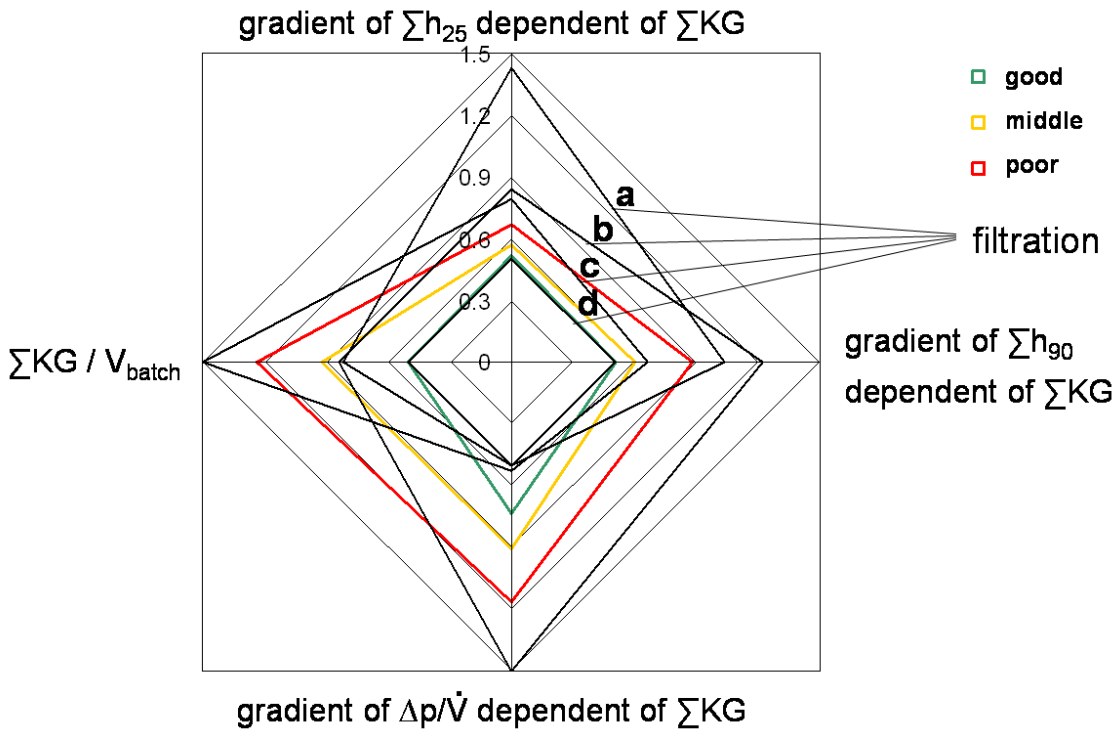


Figure 4.17: Example filtration (a, b, c and d) graphically represented with web diagram

as centroids for the clustering.

Appendix C contains the results of the clustering approaches. The results of *AGNES* are characterised by are very high agglomeration coefficient $ac = 0.99$. Figure C.3 shows the clustering of *hell A* filtration without respect to filter state. Additionally, Figure C.4 displays the clustering of the same data, supplemented by three dummy filtration (*good*, *middle* and *poor*). Comparison of the two illustrations shows, that the additional dummy filtration have only little impact within the clusters but do not affect the global result of Cluster Analysis.

Furthermore, an analysis using the *k-means* algorithm is carried out. As this method generates clusters around given centroids or chooses those centroids randomly from the given data, both approaches are applied. In Appendix, figures C.5 and C.6 display the complete results, Table 4.7 and 4.8 summarise these clusterings.

Obviously, the algorithm determines the same cluster using random centroids and using dummy filtration as centroids, only the sequence of clusters differs.

Table 4.7: Summary of *k-means* results using random centroids

cluster	**elements**	$\bar{b}_{h_{25}}$	$\bar{b}_{h_{90}}$	$\bar{b}_{\Delta p/\dot{V}}$
1	34	0.129	0.177	0.915
2	6	0.156	0.230	2.802
3	134	0.141	0.205	0.264

Table 4.8: Summary of *k-means* results using dummy filtration as centroids

cluster	**elements**	$\bar{b}_{h_{25}}$	$\bar{b}_{h_{90}}$	$\bar{b}_{\Delta p/\dot{V}}$
1	134	0.141	0.205	0.264
2	34	0.129	0.177	0.915
3	6	0.156	0.230	2.802

The cluster means of slopes $\bar{b}$ and the number of elements within the three clusters equate.

Furthermore, the clusters found by *AGNES* and *k-means* are very similar, assuming a splitting of the clusters given by *AGNES* at the level yielding the same number of clusters than *k-means*. Comparing the results of both methods (s. Appendix C, fig. C.3 and fig. C.4) shows, that the same small group of six elements appears using the hierarchical method, marked by a high dissimilarity to the other elements. The separation of the remaining two cluster differs exiguous from *AGNES* to *k-means* by ten elements.

4.2.4 Evaluation of Clusters

Adjacent, the unlabelled groups are evaluated. Because of the features of this approach, the clusters can be evaluated regarding two property. Firstly, dummy filtration enable evaluation of clusters as they act as tracers among the regular filtration. Table 4.9 describes the used dummies.

Thus, by this means the cluster can be labelled. The already presented results of *AGNES* and *k-means* allocate the dummies differently. Both Methods yield a large cluster (*AGNES*: 124 elements, the left branch; *k-means*: 134 elements) comprehending the *good* dummy and another group (*AGNES*: 44 elements, the middle branch; *k-means*: 34 elements) containing the *poor* dummy. The third dummy filtration (*middle*) resides among those 10 filtra-

Table 4.9: Description of dummy filtration

description	**filtration id**	$b_{h_{25}}$	$b_{h_{90}}$	$b_{\Delta p/\dot{V}}$
good	10000	0.046	0.093	0.257
middle	10040	0.069	0.139	0.482
poor	10080	0.083	0.222	0.617

tion assigned to the large cluster by *k-means* but to the medial cluster by the *AGNES* method, respectively. For both the hierarchical as the partitioning method, the third cluster, having six elements, can not be labelled using a dummy.

Therefore, the evaluation is carried out with the help of cluster means as summarised in tables 4.7 and 4.8 in comparison with the slopes of dummy filtration (Tab. 4.9). As the remaining cluster is characterised by attributes being notedly higher than those of the *poor* dummy with particularly an outstanding high slope $b_{\Delta p/\dot{V}}$, this group is identified as a group of filtration showing *very poor* characteristics compared to dummy filtration.

4.2.5 Interpretation of Clusters With Decision Trees

The results of Cluster Analysis are further analysed by means of Decision Tree algorithm, as already used before. Target parameter for construction of Decision Trees is given by the clustering, by the number of the clusters. The Appendix contains figures C.7 and C.8, showing the resulting trees, both based on *AGNES* and *k-means* clustering.

In this case, the Decision Trees are completely based on the performance attribute $b_{\Delta p/\dot{V}}$. Splitting of the data universe using this attribute yields trees with nearly pure leave nodes.

4.3 Knowledge Discovery

This section takes a deeper look at the input data, the laboratory data. These data are analysed using three methods: the Decision Tree method, a method based on the Expectation Maximisation algorithm and the Principal Component Analysis. The results are presented in the following.

This knowledge discovery aims at revealing patterns, associations or structure within the data which enhances the prognosis task, but furthermore it can be basis for case studies. As already mentioned above, storage time T_s, original gravity G_o, alcohol content, both per weightA_w and volume A_v, apparent E_s and real E_w extract, apparent V_s and real V_w attenuation, vicinal dicetones DA, pH-value and colour F are available for these analysis.

4.3.1 Decision Tree Analysis of Laboratory Data

An first overview on input data is gained by a Decision Tree analysis of laboratory data using filtration classes as target attribute for construction of the tree. All of the laboratory attributes described above are used as input data. The complete resulting Decision Tree for the example of beer *lager A* is presented in Appendix D (Fig. D.1). The most important attributes for tree construction are storage time, alcohol content, attenuation, pH-value and colour. Original gravity appears only on some leafs at the bottom of the tree. The impact of storage time stands out among the other attributes. The insight this Decision Tree gives on laboratory datasuffices to extract information for construction of fuzzy sets and indicates influences of the storage conditions on the filtration. In the following, further investigations on storage conditions are carried out and later laboratory data is analysed in detail by means of EM algorithm and PCA.

Figure 4.18 shows the same tree with a coloured bar representing the classification (from dark red as *very poor* to dark green as *very good*) of the underlying filtration. The leafs show groups of good homogeneity. Furthermore, all nodes described by the outstanding attribute T_s are marked red.

Obviously, this figure marks the role storage time T_s has. Comparison of the splitting of the data caused by storage time and the corresponding classifications indicates an important influence of T_s on filtration.

Following this indication, further investigations on influence of storage time are carried out. Therefore, the distribution of *better* and *poorer* filtration with respect to storage time is analysed. Figure D.2 shows these distributions of *lager A* and *lager B* separately for first and second filtration.

Figure 4.19 displays this disturbion of filtration classes regarding storage time in detail. It reveals a storage time $T_{s,\max}$ which splits the storage time in two ranges. The range having longe storage time $T_s > T_{s,\max}$ shows a poorer rate of *better* and *poorer* filtration classes than filtration with shorter storage time. An optimal range characterised by an advantageous rate of

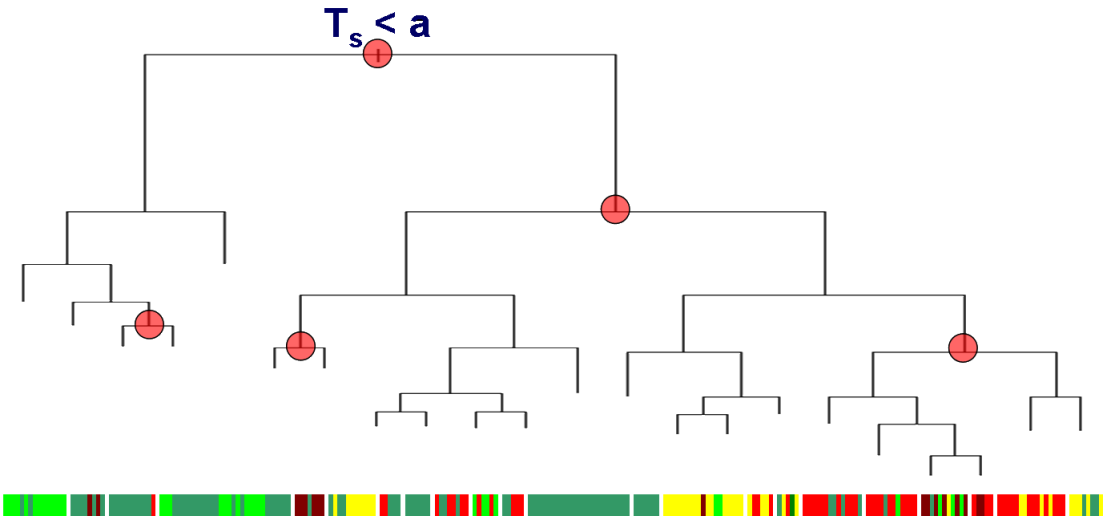


Figure 4.18: Decision tree of laboratory data

filtration classes is limited at the lower bound by $T_{s,\min}$, due to technological requirements like reduction of vicinal dicetones.

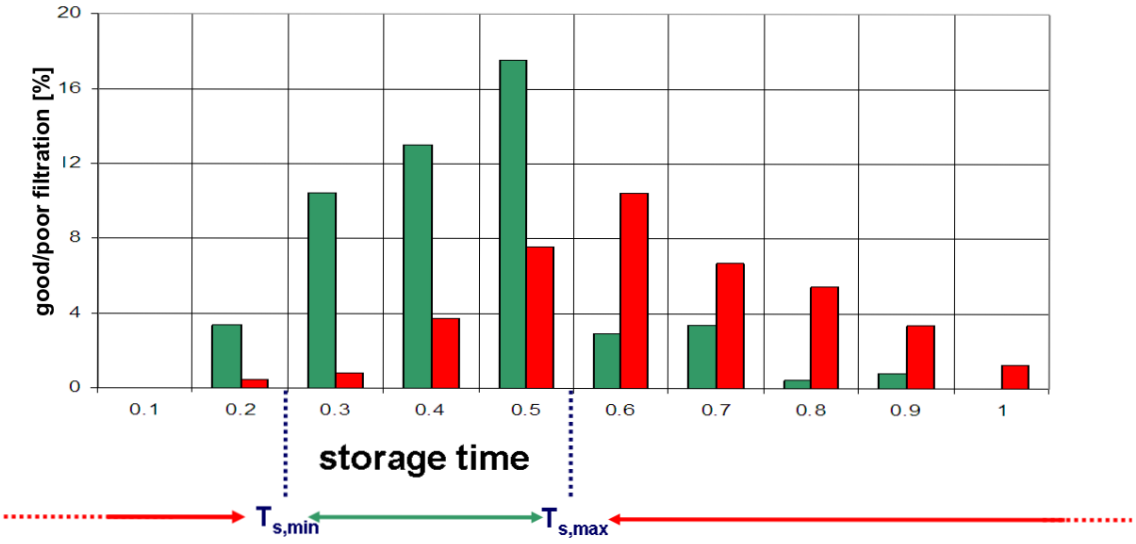


Figure 4.19: Distribution of filtration classes regarding storage time for the example of *lager A* first filtration

As storage time is one feature in the field of storage conditions, another is the geometry of the storage tank. In breweries, horizontal tanks were commonly

used for years, but these days vertical tanks are common. With the geometry of the vessel also the storage conditions like for example pressure differ. The beer type *lager B* permits the consideration of geometry as it is stored in both types of vessel with major utilisation of the standing vertical tanks. The distribution of filtration classes with respect to tank geometry is presented in figure 4.20.

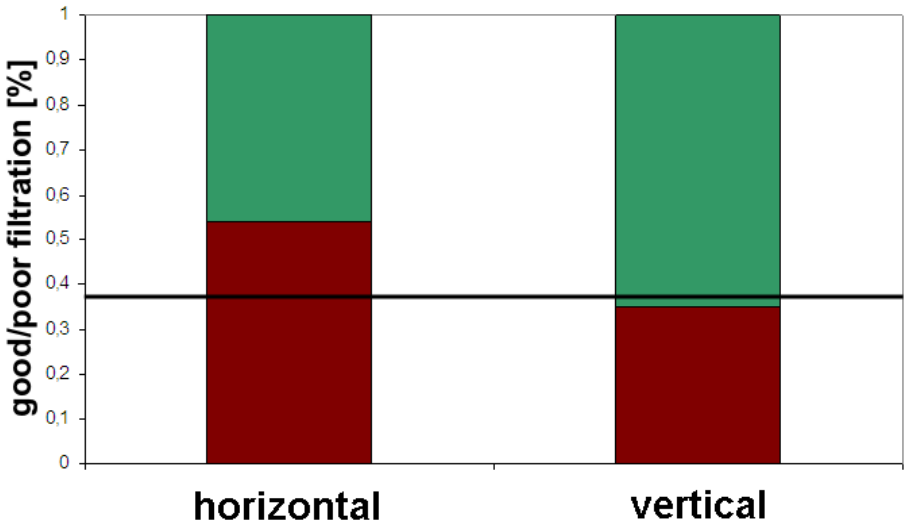


Figure 4.20: Influence of storage tank geometry on distribution of filtration classes

The black horizontal line marks the ratio of *better* to *poorer* filtration of all filtration of this beer type *lager B*. The ratio of only filtration stored in vertical tanks is very similar, but the ration considering horizontal tanks is adverse. The low number of filtration stored in horizontal tanks $n_h = 24$ compared to $n_v = 163$ vertical stored filtration attenuates the result, but still it indicates a marginal influence of geometry.

The last feature in the field of storage conditions examined is storage department. The storage cellar is split up in several different departments, each containing several tanks. Each department is equipped with its own lie system and blending device. The beer type *lager A* is stored in eighteen different departments all consisting of horizontal tanks of identical size. Thus, investigation on influence of storage department on filtration is carried out for this beer.

Regarding the first classification approach this beer type shows a ratio of *better* to *poorer* filtration $r_{better/poorer} = 0.5$. Assuming no influence of the source department on filtration classification the distribution of filtration

classes with respect to departments will show the same ratio for all departments with minor variations. Therefore, this distribution is presented in figure 4.21.

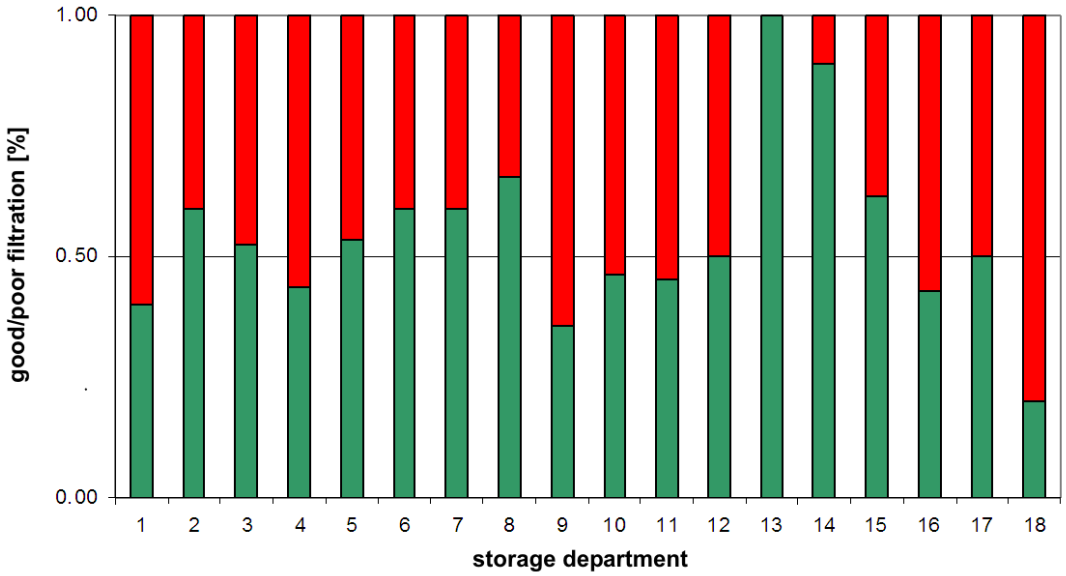


Figure 4.21: Influence of storage department on filtration classes

With exception of three departments (Dep. 13, 14 and 18) all other departments show only the expected fluctuations of distribution of filtration classes. Departments Dep. 13 and Dep. 14 attract attention by notedly better ratios of *good* to *poor* filtration, whereas Dep. 18 is identified having an unfavourable ratio.

Alltogether, from storage conditions only storage time shows a significant impact on filtration, whereas storage departments and the geometry of tanks affect filtration only marginal.

4.3.2 Analysis of Laboratory Data with EM Algorithm

Within the presented thesis, the Expectation Maximisation (EM) algorithm is used for analysis of laboratory data, to reveal structures within these data. Therefore, EM algorithm is applied on each laboratory attribute of each beer type. Targets of analysis are the frequency distributions of laboratoy data as well as filtration data.

Firstly the ranges for the distribution of frequency of occurrence are determined. Starting from a maximum of 100 ranges with a very small width the number of ranges is decreased iteratively until the ranges are representative. As a criterium for further expansion of range width the number of ranges containing not more than 3 elements has to be less than one third of all ranges. Figure 4.22 shows a frequency distribution using to different range width. The absolute frequency is presented over the ranges. Therefore, the scale of abscissae changes. The ranges used for part **b** of this figure is three times wider than the range applied at part **a**. Thus, the noise of the distribution is reduced.

The next step, assigning of superposed distributions, an expectation of those distributions has to be made. This is carried out by identifying the peaks of possible distributions. Thus, as shown in figure 4.23, the maximum values, marked by the red arrows, are determinded, together with their right side limit value, the next peak minimum, visualised by the green arrows. For each possible distribution, these two values, supplemented by an assumed standard deviation $\sigma = 1$ are further processed. If, like for the left peak, no possible minimum can by identified only the peak maximum is considered, marked by both a red and green arrow.

Subsequently, these values, used as a first expectation, are processed with the help of the Microsoft Excel Solver to compile subordinate distributions. Therefore, by the means of EM algorithm, those distributions are calculated iteratively. Figure 4.24 shows the results of an example distribution. The coloured lines represent the subordinate distributions.

The exemplary distributions of laboratory data attributes alcoholic content A_w, real extract E_w and original gravity G_o, both for *lager A* and *lager B*, are presented in Appendix D, figures D.3-D.6.

The number of subordinate distributions differs for the attributes, but the figures show very similar underlying distributions for the same attribute of those two different beer types. Thus, the original distribution of for example alcoholic content (Fig. D.3 and D.4) is represented by five subordinate ones. The ranges of these distributions are used for a gradation of the attribute. For other attributes like original extract and attenuation more subordinate distributions describe the frequency distribution.

As a further simplification those peaks are conglomerated to 5 classes identified as *lower*, *low*, *middle*, *high* and *higher*. Therefore, the three most important peaks are considered: one of these peaks represents the peak of the overlying distribution, the two others are chosen among the left and right neighbour peaks, still overlapping the first one. These peaks, from left to

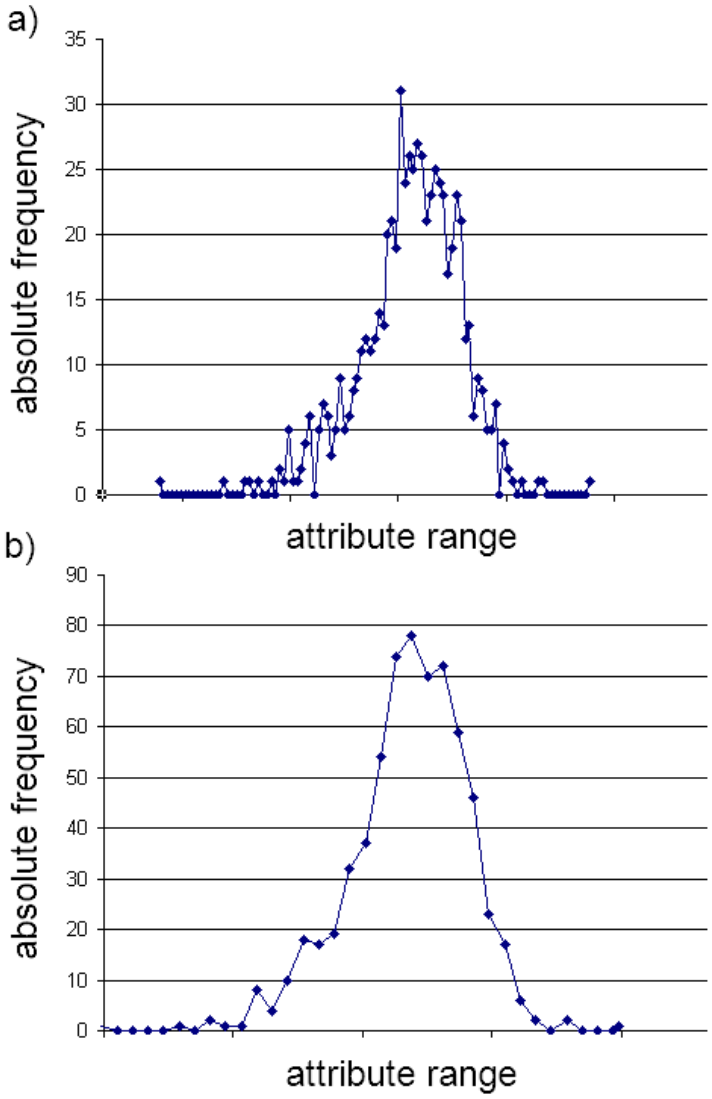


Figure 4.22: Example of a frequency distribution using small (a) and wide (b) ranges

right are identified as *low*, *middle* and *high*. The range left of *low* is marked as *lower*, the *higher* class is defined as right of *high*.

Analogue, classification of filtration attributes like slopes of haze values and filter resistance is carried out. Those resulting classes of filtration and lab-

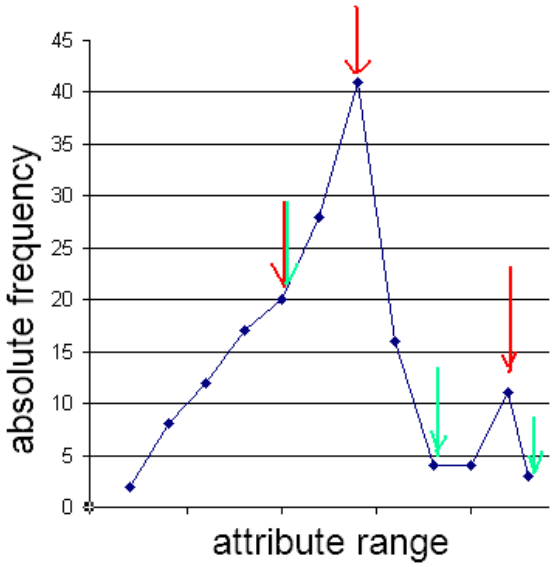


Figure 4.23: Example of identification of possible superposed distribution

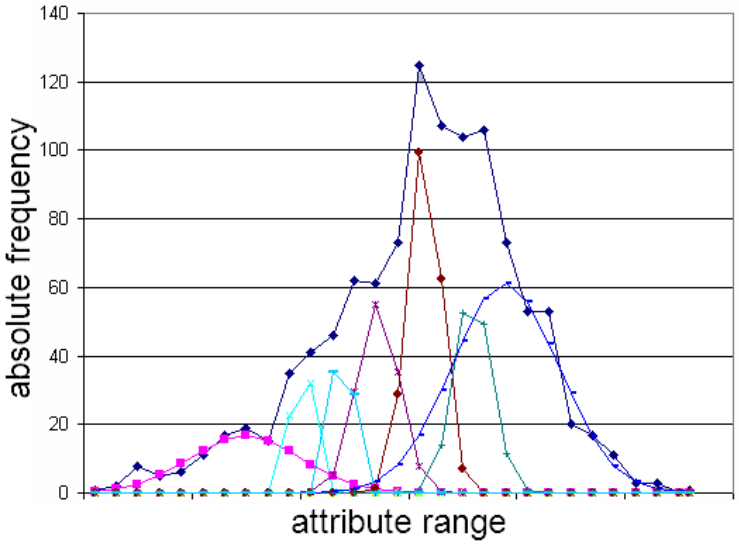


Figure 4.24: Example of superposed distributions

oratory attributes and their limitating values are for example applied for defining fuzzy sets.

4.3.3 Principal Component Analysis of Laboratoy Data

As a further analysis of laboratory data and specially the influence on filtration a Principal Component Analysis of laboratory data is carried out. Input parameters of this approach are the laboratory attributes as already used before: G_o, A_w, E_w, pH, C, DA and t_S. They are amended by the EM classification of slope $b_{\Delta p/\dot{V}}$ as target attribute.

The PCA transforms the attributes to principal components. The following table presents the cumulated percentages of principal components in total variation.

Table 4.10: Cumulated percentages of PCA of laboratory data

component	pc_1	pc_2	pc_3	pc_4	pc_5	pc_6	pc_7	pc_8	pc_9
percentage	24.62	42.81	59.04	72.02	80.96	88.30	95.14	99.99	100.00

In the following, figure 4.25 shows the results of that analysis, presented as a biplot of the components pc_2 and pc_3. The target attribute value of the elements, the classification of $b_{\Delta p/\dot{V}}$ according to the EM algorithm, is marked by "1" for *high*, "0" for *low* and an asterisk "$*$" for *middle* slopes. The contributions of the original attributes are visualised as arrows, as described in section 3.4.6. Biplots of the remaining combinations are presented in Appendix D.

This figures demonstrates covariance of attributes. For example the nearly perfect negative correlation of storage time t_s and vicinal dicetones DA, technologically caused by degradation of DA during storaging. Moreover, this figure visualises a separation of *high* and *low* slopes of $\Delta p/\dot{V}$, but still with some overlappings. The elements characterised by *middle* slopes are distributed among both other groups. Therefore, the differences of filter resistance are not explained by this PCA.

To implement more information, two other attributes are recruited. These are batch number and the initial filter resistance of the filtration $c_{\Delta p/\dot{V}}$. Both attributes give information on the initial conditions of the filtration. Together with laboratory data the carried out PCA yields better results in separating the target attribute, as figure 4.26 displays. The target attribute values are marked within the figure as "1" for *high* and "0" for *low*. The elements showing *middle* slopes of filter resistance are identified by an asterisk($*$).

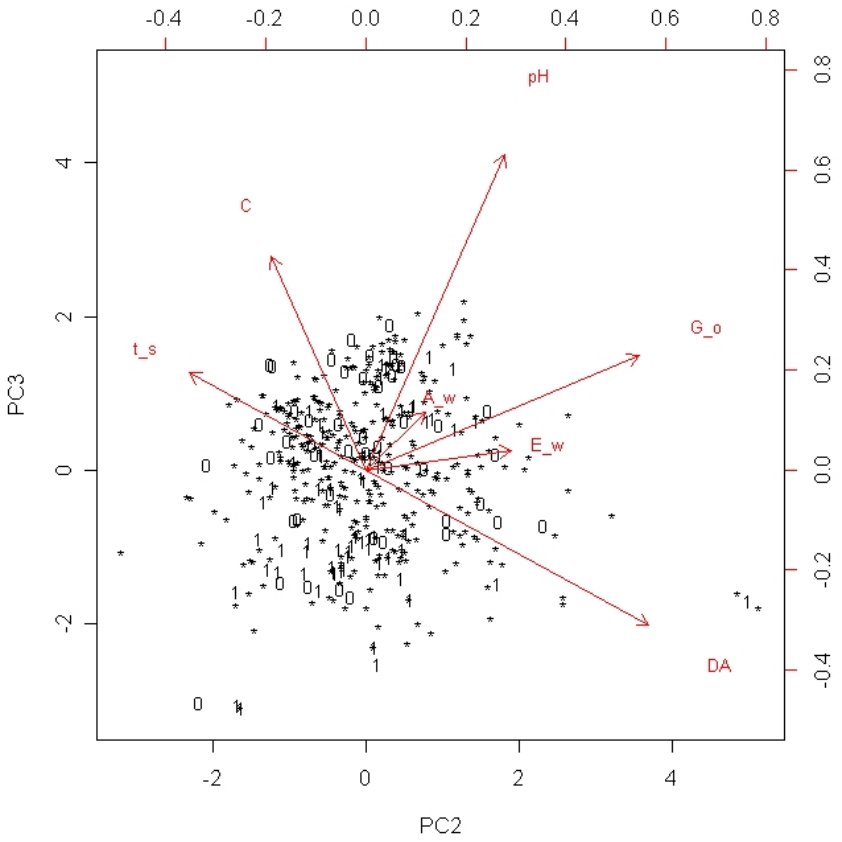


Figure 4.25: Principal Component Analysis of laboratory data regarding slope of filter resistance

Still, the technological covariance can be found, but due to the recruited attributes a better separation is yielded. Those two attributes are found to have a high correlation and influence the better separation mainly. The "noise" caused by the *middle* slopes of filter resistance still exists, but by this means the *low* and *high* ones are separated nearly completed from each other with only few overlappings.

4.4 Prediction of Filtration

Approaches in predicting filtration based on the available laboratory data are carried out using different input data (partly results from other methods),

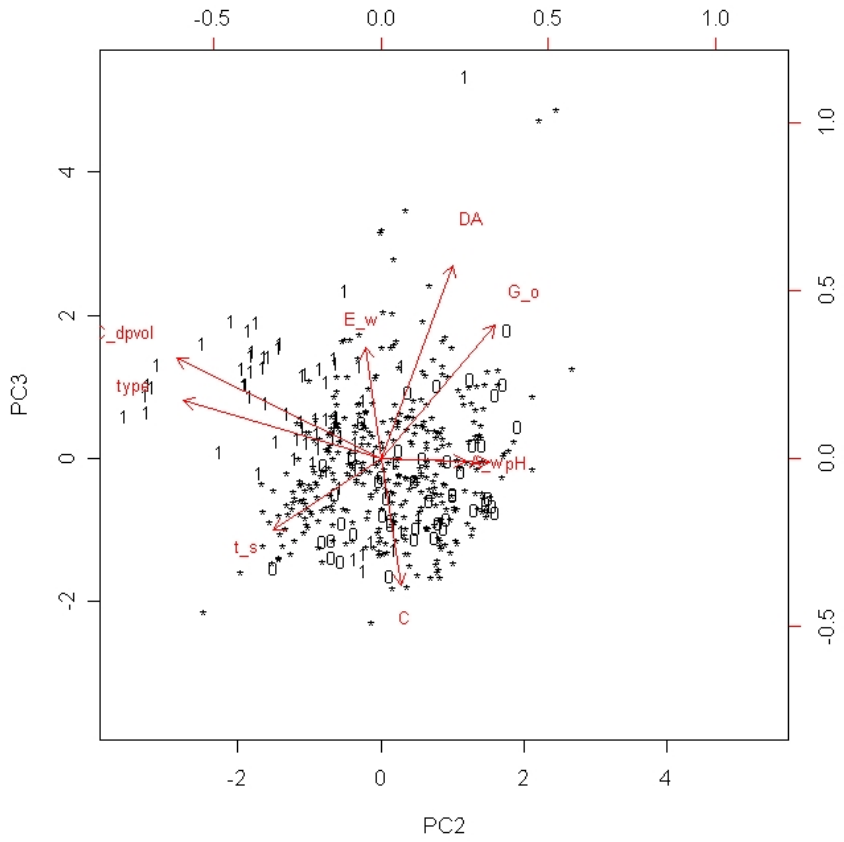


Figure 4.26: Principal Component Analysis of laboratory data supplemented by batch number and initial filter resistance regarding slope of filter resistance

different methods and also combination of methods. Even the output of those models differs from predicting tendencies of filtration (*better* and *poorer*) up to the forecasting of attribute values like the slope of increasing filter resistance ($b_{\Delta p/\dot{V}}$). These attempts and their results are described in the following.

4.4.1 Fuzzy Logic

Among the prediction methods, Fuzzy Logic approaches are established for prediction tasks. Thus, several Fuzzy Logic models are set up within this thesis. These approaches differ in the basis of fuzzy sets, both input and output data. Commonly for all approaches, the fuzzy sets are defined man-

ually from the used data, for example Decision Tree results, findings of EM algorithm or PCA, or combination of these. The respective approaches are presented in the remainder of this section.

Fuzzy Sets derived from Decision Tree of laboratory data and first classification

Based on the results of Decision Tree analysis of the laboratory data the fuzzy sets are defined. Therefore, the limit values of each attribute yielded by the tree are accounted together with the respective classes of corresponding filtration by means of *better* and *poorer* filtration.

The resulting fuzzy set yields results of maximal 60% correct predictions, even after adapting of the set.

Fuzzy Sets derived from Decision Tree results and EM classes of filter resistance

For the approach, the input sets are also constructed based on the Decision Tree of laboratory data. Differing to the first approach, this time target attribute is the slope of filter resistance and its classes generated by EM algorithm. The resulting fuzzy set is very lean as it contains only 14 rules. It shows results of up to 65% correct predictions.

Fuzzy Sets derived from EM algorithm results

This third approach uses the results of EM algorithm of laboratory to construct the input sets of the fuzzy system. Figure 4.27 shows the input set of *stw* as an example.

Also, the output set base upon EM results, respectively of those from the analysis of slopes, especially in this case the slope of filter resistance. This proceeding offers results of about 50% up to 60% at maximum, depending on the defined fuzzy sets.

Fuzzy Sets derived from various results

This final Fuzzy Logic approach combins the rules of the approach based on Decision Tree findings and those based upon EM results. This proceder leads to results of maximal 70% of correct predictions.

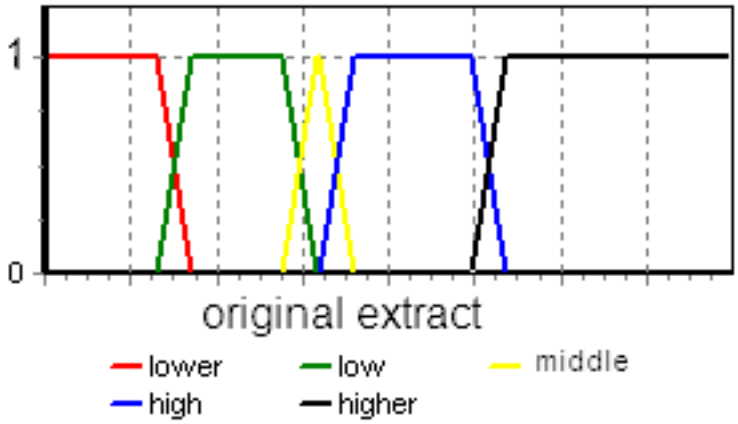


Figure 4.27: Fuzzy input set of original extract *stw*

4.4.2 Artificial Neural Networks

Furthermore, within this work ANNs are applied for prediction of filtration. Here, two approaches are carried out, predicting firstly the classes resulting from the first classification approach and secondly those of the second approach. For each approach network topology of the ANN as well as the learning algorithm are evaluated. Overall, the best results for those approaches are reached by a net of six input nodes, one hidden layer having four nodes and at last two output nodes. Resilient propagation qualifies best as training algorithm. Figure 4.28 displays the applied topology for the example of the classes resulting from the first classification approach.

ANN based on laboratory data for prediction of *better* and *poorer* filtration according to the first classification approach

After evaluation of different topologies reaching from up to four layers with overall 25 nodes down to networks having three layers with only 12 nodes, best results are reached applying the topology presented above. This ANN yields correct predictions for about 70% of the testing data.

ANN based on laboratory data for prediction of *very good* and *very poor* filtration according to the second classification approach

The topology as well as the results of this second ANN approach are very similar to those of the first ANN approach. The difference in topology is the

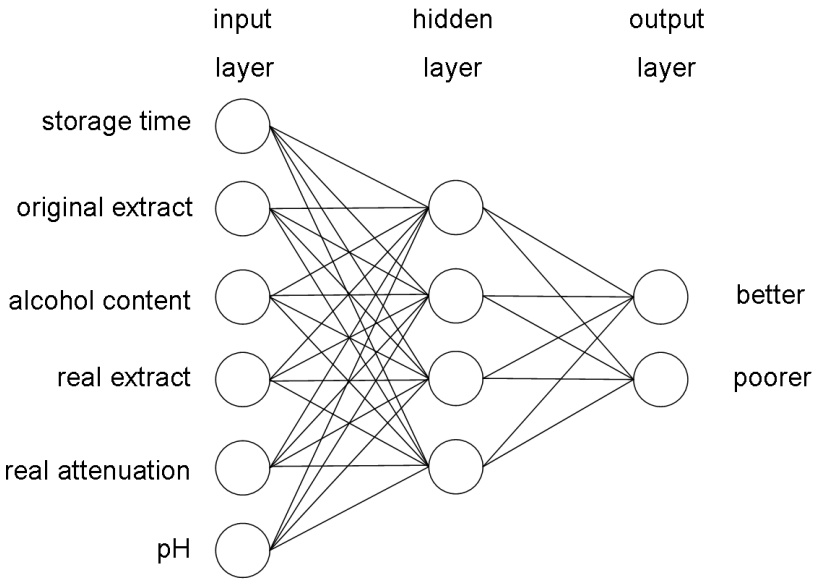


Figure 4.28: Topology of applied ANN

output nodes representing the *very good* and *very poor* filtration. Again, the ANN results a prediction rate of about 70%.

4.4.3 Predict Tree

This method uses a Decision Tree for prediction. Therefore a Decision Tree is constructed by a subset of data, the training data. The presented method predicts the target values of the remaining testing data exerting that Decision Tree. In this work several approaches based on this method are carried out, differing in input data and target values.

Predicting *better* and *poorer* filtration according to the first classification approach

For this approach, the Decision Tree is set up from laboratory data. The classes *better* and *poorer*, yielded by the first classification approach, are applied as target values as well as the rating from "0" to "5". The Predict Tree yields results of about 50% of the precise rating and up to 70% for the first mentioned two classes.

A deeper look on the generated trees indicates some inhomogeneous leafs, representing not only elements of contiguous ratings but of distant ones. Considering these groups as negative influences factors on prediction, the classification carried out is adapted by classifying only homogenous classes as described earlier and identifying and labelling the inhomogeneous ones. Thus, the original ratings $0 \ldots 1$ as well as the derivated classes *better* and *poorer* are complemented by the class *inhomogeneous*. This reclassification improves prediction quality up to 70% for prediction of rating and to about 80% for coarse classes.

Predicting *very good* and *very poor* filtration according to the second classification approach

The classification of the second approach together with laboratory data generates the Predict Tree for this approach. The prediction of those classes of *very good* and *very poor* filtration succeeds for about 70% of all data. Like for prediction of the classes of the first classification approach, this prediction shows false postives and false negatives.

Predicting the slope of filter resistance form laboratory data

The third approach carried out with this method predicts the slope of filter resistance $b_{\Delta p/\dot{V}}$. As a precise prediction of the real value is very inprobable, a distance $d = |b^{real}_{\Delta p/\dot{V}} - b^{predict}_{\Delta p/\dot{V}}|$ of predicted and real value is applied. As the values are normalised to the range $0 \ldots 1$ the maximum distance $d_{\max} = 1$. Figure 4.29 displays the results of prediction dependent of this distance d.

The green line shows the increase of prediction quality with increasing distance d. At a distance of $d = 0.24$ a prediction rate of more than 80% is reached. The false prediction decrease from 64% higher and 33% lower predictions down to 7% and 11% for a distance of $d = 0.24$, as the yellow line for lower and the red for higher predictions shows.

Predicting the slope of filter resistance form the laboratory attribute classes yielded by EM algorithm

This last Prediction Tree approach recruits the classes of laboratory data attributes extracted from EM algorithm analysis of laboratoy data. This causes a generalisation of input data. Output data, the target attribute,

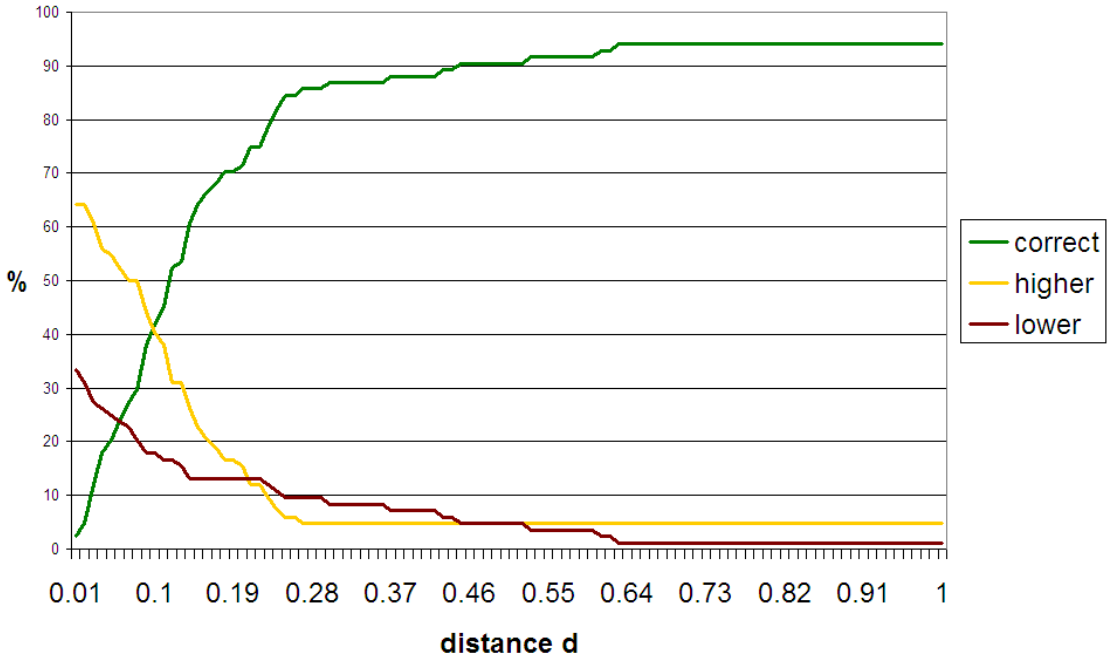


Figure 4.29: Prediction results dependent on distance d of predicted and real value

features again the slope of filter resistance $b_{\Delta p/\dot{V}}$. This approach results correct prediction in around 70% of validation cases.

4.4.4 Conclusion of Prediction Approaches

The following table 4.11 sumarises the results of prediction approaches.

Obviously, the best results show prediction rates of 80% at maximum. The following Chapter 5 discusses these results and their preceding process steps in the Data Mining process.

Table 4.11: Summary of prediction approaches and results

method	**input data**	**target attribute**	**prediction result**
Fuzzy Logic	Decision Tree of laboratory data	classes *better* and *poorer*	$\sim 60\%$
Fuzzy Logic	Decision Tree of laboratory data	EM classes of $b_{\Delta p/\dot{V}}$	$\sim 65\%$
Fuzzy Logic	EM classes of laboratory data	EM classes of $b_{\Delta p/\dot{V}}$	$\sim 60\%$
Fuzzy Logic	combination of EM classes and Decision Tree of laboratory data	EM classes of $b_{\Delta p/\dot{V}}$	$\sim 70\%$
ANN	laboratory data	classes *better* and *poorer*	$\sim 70\%$
ANN	laboratory data	classes *very good* and *very poor*	$\sim 70\%$
Predict Tree	laboratory data	classes *better* and *poorer*	$\sim 80\%$
Predict Tree	laboratory data	classes *very good* and *very poor*	$\sim 70\%$
Predict Tree	laboratory data	$b_{\Delta p/\dot{V}}$	$< 80\%$
Predict Tree	EM classes of laboratory data	EM classes of $b_{\Delta p/\dot{V}}$	$\sim 70\%$

Chapter 5

Discussion

This chapter discusses the methods and results of the presented thesis. Considering the presented CRISP-DM process model for Data Mining tasks (Fig. 3.1) a constant evaluation and discussion of methods and results is necessary. Due to the formal structure of this thesis this repeating proceeding is presented in a linear way in the following.

In the following the three previous defined subtasks *classification*, *knowledge discovery from laboratory data* and *prediction* are discussed. On top of that a conclusion of the complete Data Mining process, its applicability and its assets and drawbacks is given for especially the technological motivation and target of this presented thesis.

5.1 Classification Approaches

During this work, two approaches of classifying filtration are carried out. Target of these approaches is the allocation of relevant information included within the process data. Therefore the data is reduced to manageable units.

These approaches differ mainly in the process of data reduction, but thereby the subsequent processes are bothered. This influence and the methods themselves are discussed in the following regarding *applicability* and *findings*.

5.1.1 Potential Of The Used Methods

Data Reduction

Both data reduction approaches of this work show their own excellences and impairments. The first approach enables a detailed view on each of the used attributes. Beside that, the description of each filtration by only the arithmetic mean does not allow a considering of individual characteristics. Further on, this proceeding determines examination of only the first filtration of each batch. Another impact on this method is given by technological constraints, as the kieselguhr dosage is regulated for the first filtration in a declining way, so that these filtration show high amounts of kieselguhr. This is seen especially in filtration of small volumes.

The second approach is carried out to remedy most of the restrictions mentioned above. As shown, the integration of attribute values of the filtration time series yields monotonous functions. Furthermore, a regression of these functions enables a high reduction of the process data, comparable to those reduction of the first approach. The axis intercept of those regressions describes the initial state at beginning of each filtration. Thereby, the comparability of filtration is limited only to the beer type, but no more to the position of the filtration within the corresponding batch.

Therefore the second approach applies obviously better as it enables the further processing of all filtration. In reference to the results of both classification approaches, identifying on the one hand the quality attributes (h_{25} and $_{90}$) as main influences (first approach) and on the other hand the performance attributes (Δp, $\dot{V}$ and $\sum KG$ combined by $b_{\Delta p/\dot{V}}$), each approach seems to enable a different view on the data.

Cluster Analysis

Especially within the first classification approach several clustering algorithms and their parameters were analysed. Considering their results the algorithm *AGNES*, showing the best separation of groups is chosen for further processing. This method reveals structures and patterns of fitration data and therefore forms the fundament for the adjacent classification of these filtration.

Evaluation

With respect to the two data reduction approaches, two adequate ways of evaluating clusters are carried out. The Assessment applied for the first approach is based upon an assessment of each of five filtration attributes, yielding a 5 digit binary evaluation. This may be summarised to a 1 digit decimal number, yielding a range $0 \ldots 1$, or even to a two class allocation, *better* and *poorer*. Particularly the binary evaluation suffices for a detailed look on filtration. Thus, a discussion with experts revealed information that special combinations of attribute evaluations indicate reasons. Most notably is the combination of a high value of pressure increase Δp^{ast} and high mean of haze h_{90} which may be indicating high values of β-glucanes [66].

Dummy filtration are introduced for the review of the clusters (second approach). These virtual filtration are defined with respect to the threshold values used by experts for evaluation of filtration. The presentation of dummy and real filtration within a web diagram enables the evaluation and discussion of filtration. Furthermore, these dummy filtration act as tracers within the Data Mining process and allow therefore an opportunity of evaluation.

Decision Tree

This method within this subtask *classification* reviews the evaluation of filtration clusters. Particularly for the first classification approach, this method has shown its applicability by revealing nearly the experts threshold values for filtration attributes, as described in Chapter 4 with this example of haze h_{25}.

5.1.2 Findings

The outstanding result of this subtask is given by the possibility of a machine-aided classification of filtration, based on Data Mining methods. The fact that these classes of Data Mining methods comply those classes experts would define underlines the applicability of the approach and the chosen Data Mining methods.

Another important result of this subtask is the qualification of the second classification approach which is based on a data reduction by generation of monotonous functions of process data. This leads to the conclusion that the investigation of all filtration of one beer type without further limitation gets possible.

5.1.3 Open Questions

Concluding, the previously discussed classification approach suffices according to the defined subtasks targets. Nevertheless, some open questions remain. These questions deal mainly with an enhancement of data reduction and therefore of the whole classification.

Following the discussion of data reduction, one question needs to be asked, whether the assets of both approaches of reduction can be combined (e.g. by considering results of both approach for further classification).

Considering the second approach, one may think of advancing this approach, too. Purchase for this is given by the regression of the monotonous function. Though the linear regression shows very good correlation to these functions, a consideration of different sections of such a function may reveal small changes of the gradient. These knicks of the gradient may further on lead to an enhanced classification and may reveal further technological impacts.

Furthermore, a third approach, the clustering of the time series, may suffice. Warren Liao [114] gives a survey on clustering techniques for time series data. More over, papers dealing with time series classification [96, 25] and Data Mining approaches regarding time series [120, 49] indicate the suitability of such approaches.

Independent of an enhancement of the data reduction methods a weighted contemplation of filtration attributes will better satisfy the constraints in practice. Thus, in practice the attribute h_{25} becomes less important [113]. Therefore, a detailed analysis of h_{90} in combination with Δp^* and $\dot{V}$ may lead to better and more practical results.

At last, the question arises, whether these approaches qualify for automatic regulation of filtration process. Especially the second approach of data reduction and the principle of monotonous functions supplemented by the knowledge of dummy filtration provide the opportunity to interfere with the process for an optimisation of the process.

5.2 Knowledge Discovery of Laboratory Data

This subtask of the defined Data Mining process presents the second fundament for the prediction task. It aims at finding patterns, associations or structure within laboratory data, the input data of prediction. The remainder of this section discusses the used methods and their results.

5.2.1 Applicability of Methods

All three chosen methods for this subtask have shown their own assets and weaknesses, each yielding relevant informartion for enhancement of the prognosis task and indicating further approaches of case studies.

One problem, somehow shown by all of the applied methods of this task, presents the partly very unspecific separation of corresponding filtration by the given laboratory data. This indicates an insuffiency of information content of laboratory data regarding the target of prediction.

Decision Tree

This method suffices perfectly for gaining a first overview on the data. More over this method and its result can be modified to fullfill the prediction task, as discussed later.

In this work, especially the Decision Tree of laboratory data revealed the importance of storage condition for filtration. It indicates these and leads the way to further analysis of storage conditions.

One restriction of this method is given by the inhomogeneous leaf nodes described in Chapter 3. Thus, a coarse structure of filtration by Decision Tree of laboratoy data is obvious, but a deep look reveals "noise" within the leafs of the tree. With "noise" in this case, the inhomogeneous leaf nodes are considered. They ma be explained in two ways. Firstly, the method is inadequate for this task or systematic errors influence the result. Secondly, the content of information of laboratory data does not suffice for a complete explanation and reconstruction of those filtration classes.

EM algorithm

Also this algorithm shows its applicability for the task, for example by revealing superposed distributions within the original distribution of laboratory data attribute values. Thus, this algorithm yields classifications of those attributes, which are used for prediction approaches.

A deeper look at these EM classes shows again unspecific patterns, suggesting the same reasons as discussed above for the Decision Tree method.

Principal Component Analysis

The PCA is the third method applied for analysis of input data. As well as the other methods, this one qualifies for the Data Mining task. Thus, it reveals already known dependencies of attributes and furthermore gives the opportunity for further case studies.

Even this method, as the figures within Chapter 3 show, does not qualify for a clear separation of filtration classes with only considering laboratory data. A supplementation of information, in this case the initial state of filter, defined by the axis intercept of linear regression of $\Delta p/\dot{V}$ over $\sum KG$, and the position of filtration within the batch, enhances the discrimination of filtration classes by means of laboratory data.

5.2.2 Findings

The analysis carried out within this task reveals several facts and indices, some of them already known. These known ones confirm the validity of the applied methods and the new ones give information for further analysis. Thus, the further view of the indicated influence of storage conditions revealed a remarkable impact of storage time on filtration, whereas storage department and tank geometry showed no significant influence. Reasons for this may be the identical equipment of storage departments and the good practice in production. The analysis of influence of tank geometry is limited, as only one beer type owns enough relevant data for such an analysis, and even for this beer type the horizontal vessel are only considered to avoid production bottlenecks.

As further important findings of this approach the unspecific allocation of filtration classes as well as the laboratory data and its structures indicate a leak of information. Thereby, the *Open Questions* arise, as described in the following.

5.2.3 Open Questions

Indeed, as discussed with both the methods and the findings, the question arises whether the laboratory data still contains information that the applied methods do not reveal or even if the methods are not applicable for this task. The third explanation for these unspecific patterns is given by a leak of information content of the input data.

Up to this point and for this application, the Decision Tree method as well as the principal component analysis are modified to their optimum. A better adaptation of the EM algorithm seems to be possible. By means of this, an analysis of those unspecific allocations and perhaps an enhancement of this allocation may perform better.

5.3 Prediction of Filtration

This last subtask combines the findings and partly the methods of the first two subtask. The results show a maximum prediction rate of about 80%. This finding and the accordant methods are discussed further on.

5.3.1 Applicability of Methods

For this subtask, three different Data Mining techniques, suited for prediction tasks, are chosen. More over not only the original input data is applied for prediction, but also the findings gained by the knowledge discovery process, such as rules and threshold values from Decision Tree old classes of laboratory data yielded by EM algorithm. The methods and the input data for prediction are considered within the following sections.

Fuzzy Logic

The fuzzy sets generated within this work are all constructed manually, by combining experts knowledge and findings of knowledge discovery process. Therefore, the results from Decision Tree analysis of laboratory data, the classes of laboratory data defined by EM algorithm and combinations of both are applied.

The first two Fuzzy Logic approaches carried out recruit Decision Tree results for prediction, but differ in the prognosed attribute. The comparison of both the prediction of the coarse classes of *better* and *poorer* filtration shows a poorer rate (60%) than the prognosis of classes of gradient of $\Delta p/\dot{V}$ (65%). This indicates the important *middle* filtration class. The two class system *better* and *poorer* prevents a correct assignment of these filtration. In opposition to this approach the EM classes include these classes, so that this second approach improves the prediction.

Furthermore, this leads to using the EM classes of laboratory data as in-

put data for construction of fuzzy sets. The results of that approach (60% prediction rate) indicate to much generalisation of laboratory data. Further investigations of the results of prediction based on Decision Tree findings and EM classes show differences in the prognosis. To use these different findings a final approach based on both Decision Tree rules and EM classes is carried out, yielding the best prediction rate of Fuzzy Logic approaches of 70%.

Artificial Neural Networks

Two approaches prognosis based on ANNs are carried out. Both approaches are based upon laboratory data as input data, one predicting the classes *better* and *poorer*, the other the extreme classes *very good* and *very poor*. Both yield a prediction rate of about 70%.

Further approaches are not considered. Especially modification of input data is obsolete, as the ANNs deal best with the laboratory data. Further investigations regarding more differing topologies and prediction attributes showed no improvement of the results. This indicates again the hypothesis that the information available from laboratory data does not suffice for a better prediction of filtration.

Predict Tree

This method, based on Decision Tree algorithm, yields the best results in predicting filtration. The presented results show that this method suffices for prediction of different classes (*better* and *poorer*) as well as prediction of the gradient. But these approaches show a non negligible rate of false positives and false negatives. This problem occurs for all of the carried out approaches. Thus, this effect is independent of the applied kind of input data as well as the predicted classes or attribute.

5.3.2 Findings

The shown and discussed prediction approaches yield results of 70-80% reliable correct prediction rate. Obviously, this does not satisfy the requirements of practice.

All approaches of enhancing those results, including modification of methods, input data and the predicted attribute or classification, showed no further increase of the prediction rate or quality of prediction.

This leads to the final conclusion that the laboratory data is to unspecific and contains not enough information for a better prediction of filtration than the presented 70-80% prediction rate.

Further on, this leads to the next section, the open questions.

5.3.3 Open Questions

The outstanding question is, how the prediction result can be further increased. The discussion of all three subtasks displays, that further modifications of the data and method of this work will not lead to further success. Therefore, an expansion of the input data pool is necessary. This raise the questions which data will yield the desired success, how the data can be expanded and by which cost.

Considering the principal idea of applying only routinely measured and stored data, the data from fermentation may by applied. Especially the state of the yeast, the laboratory data and the characteristic of fermentation may contain further valuable information. Further on, an expansion of input data down to laboratory data of raw material and process data of the whole brewing process is possible. Thus, problems will appear not by the growing data but by allocation and identification of the corresponding data for filtration.

5.4 Conclusion

With respect to the target of this work, the different approaches and the fundamental process model of CRISP-DM, some final conclusions can be drawn. Thus, this work shows that Data Mining tasks dependent not only on algorithm but that the acquiring, understanding and preprocessing of data is an important part of Data Mining. Especially for brewing process, data acquiration and preprocessing is very difficult. More over, the correct identification of for example the linkage of storage tanks and filtration or fermentation and storage tank have important roles. Thus, the Data Mining process can be carried out with data distributed over several PDAs and storage systems, but a well organised data warehouse, like modern process control systems offer, alleviates the Data Mining process.

Some other aspects become apparent with this work. The process modell for the Data Mining task, in this case the CRISP-DM, guarantees usefull results. Furthermore, the role of business and data knowledge has high impact on the

Data Mining process. It is very important to understand the brewing process and its problems and its procedures as well as the used data to implement the Data Mining process.

Chapter 6

Summary and Outlook

This present thesis deals with the technological problem of predicting filtration, as filtration of beer is an important process step within the brewing process. Thus, this work applies Data Mining methods for the prediction task of filtration. Data Mining is also denoted as the process of Knowledge Discovery in Databases (KDD). The fundamental database for the prediction task of this thesis is given by a Production Data Acquisition System (PDA) of a Bavarian brewery. The data of the past five years is analysed by different Data Mining methods. Those Data Mining methods are already established for modelling target attributes in economy or finance. First approaches of adopting Data Mining methods to process technology are carried out latterly. The main objective of this thesis is to adopt Data Mining methods to develop a model for prediction of filtration. The input data of this model is constrained to routinely acquired laboratory data of matured beer and the corresponding filtration data. Especially the limitation of taking only routinely taken in consideration alleviates this approach. A consideration of non routinely taken data increases the effort and moreover reduces the available data, as the access to past data is very difficult. Regarding data of preceding process steps than the laboratory data of matured beer complicates the analysis as allocation to the corresponding filtration is very complex.

To fulfil the prediction task the considered two parts of data, in a first step filtration data and laboratory data are analysed separately. The filtration data represented by haze, volume flow, kieselguhr dosage and increase of pressure difference are classified. Therefore, two different approaches are carried out, evolving from different assumptions. One approach bases on a very simple view on filtration which is regarding the means of filtration attributes. The other one respects more details of filtration characteristics.

This is reached by transforming the filtration attributes and their time series into monotonous functions by means of summation or the consideration of dependencies of attributes with respect to technological facts. Besides the advantages and disadvantages of those two approaches the applied Data Mining methods Cluster Analysis and Decision Trees showed their aptitude for this classification task. The results of classification compared to experts knowledge and thresholds used in practice support the determined classes and their class limits, as found by the Data Mining methods.

Furthermore, the laboratory data is studied to find knowledge in form of dependencies or associations which contribute to the prediction task. By applying several Data Mining methods including Decision Trees, Expectation Maximisation algorithms and Principal Component Analysis this subtask is carried out. These investigations revealed several facts, of which some are already known. These known facts show again the aptitude of the Data Mining methods. One main new finding of this knowledge discovering process is the revealed influence of storage conditions. Further studies on this topic yield an important impact of storage time on filtration. An optimal storage duration according to filtration is found for the studied beer.

Adjacent, the results of these two parts, the classification and the knowledge discovery, are combined for the main task, the prediction of filtration based on laboratory data. For this objective of the present thesis Fuzzy Logic, Artificial Neural Networks and Prediction Trees, a modification of the Decision Tree algorithm, are applied. The resulting models predict filtration with a success rate of up to 80%. This prediction rate does not suffices the requirements for practical implementation.

Due to the following equation already presented in Chapter 3.6,

$$Results = Model + Findings, \tag{6.1}$$

the results of this approach are not constraint to models. The important finding of the prediction task indicates a lack of information. Up to a certain degree prediction suffices independent of the applied methods. This prediction rate of about 80% can not be enhanced by means of methods. Therefore, it has to be considered by which measures or attributes the database has to be extended, to suffices the requirements of practice.

Possible data which is not to complex to be included is the data of fermentation. Especially the state of the used yeast is recommended by experts as well as the temperature conditions during fermentation and maturing. But as these data are not stored in a PDA, they have to be acquired manually.

Recapitulating, this thesis shows the aptitude of Data Mining methods for

classification, knowledge discovery and prediction task for technological processes in breweries. Different models are generated enabling a prediction rate of up to 80%. Further enhancement of prediction is inhibited by a lack of information available from the applied laboratory data.

Chapter 7

References

Bibliography

[1] Adriaans, P., Zantinge, D. (1996). Data Mining. Addison-Wesley.

[2] Anderson, J. A. (1995). An Introduction to Neural Networks. Massachusetts Institute of Technology.

[3] Annemüller, G. und Schnick, T. (2001). Entwicklung einer Modelltrübungssuspension. Brauwelt. Nr. 15/16, S. 564-569.

[4] Apté, C., Weiss, S. (1997). Data mininig decision trees an decision rules. Future Generation Computer Systems, 13, 197-210.

[5] Arnold, S., Becker, T., Delgado, A., Emde, F., Follmann, H. (2002). Fuzzy-basierte Prozessführung zur mikrobiellen Produktion von Essigsäure in hoher Konzentration. Proceedings zum Colloquium Wasser-/Abwassertechnik Nachhaltige Produktion in der Getränke und Lebensmittelindustrie.

[6] Bacon, L. D. (2002). Handbook of Data Mining and Knowledge Discovery. Oxford University Press.

[7] Bandemer, H., Gottwald, S. (1993). Einführung in Fuzzy-Methoden, Akademie Verlag.

[8] Barbará D., Chen P. (2001). Using Self-Similarity to Cluster Large Data Sets. Data Mining and Knowledge Discovery.

[9] Brdossy, A., Duckstein, L. (1995). Fuzzy Rule-Based Modeling with Applications to Geophysical, Biological and Engineering Systems. CRC Press, Boca Raton.

[10] Becker, T., Delgado, A., Schmidt, H.-L. (1997). Über das Potential kognitiver Algorithmen für die Bioprozeanalytik. DECHEMA-Jahrestagung 1997, 305-307, Münster.

[11] Becker, T., Enders, T., Delgado, A. (2002). Dynamic neural networks as a tool for the online-optimization of industrial fermentations. Bioprocess and Biosystem Engineering 24, pp. 347-354.

[12] Benning, R., Becker, T., Delgado, A. (2001). Initial studies of predicting flow fields with an ANN hybrid. Advances in Engineering Software 32, p. 895-901.

[13] Benning, R., Becker, T., Delgado, A. (2002). Principal Feasability Studies Using Neuro-Numerics for Prediction of Flow Fields. Neural Processing Letters 16, p. 1-15.

[14] Benning, R., Petermeier, H., Delgado, A., Hinrichs, J., Kulozik, U., Becker, T. (2002). Verbessertes Foulingverhalten von Röhrenwärmetauscher durch Anwendung eines hybriden Modells im Prozessdesign. Chemie Ingenieur Technik 74, S. 1492-1496.

[15] Berry, M. J. A., Linoff, G. (1997). Data Mining Techniques for Marketing, Sales and Consumer Support. John Wiley & Sons.

[16] Biewer, B. (1995). Fuzzy-Methoden - Praxisrelevante Rechenmodelle und Fuzzy-Programmiersprachen, Springer Verlag.

[17] Bishop, C. M. (1996). Neural Networks for Pattern Recognition.

[18] Bock, H. H. (1974). Automatische Klassifikation. Vandenhoeck & Ruprecht, Göttingen.

[19] Bothe, H.-H. (1995). Fuzzy Logic - Einführung in Theorie und Anwendungen, Springer Verlag.

[20] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (1984). Classification an Regression Trees. Wadsworth 1984.

[21] Brunk, C., Kohavi, R. (2002). MineSet. Handbook of Data Mining and Knowledge Discovery. Oxford University Press.

[22] Buntine, W. (1992). Learning Classification Trees. Statistics and Computing, 2, 1992, 63-73.

[23] Burl, M., Asker, L. Smyth, p., Fayyad, U., Perona, P., Crumpler, L. and Aubele, J. (1998). Learning to recognize volcanoes on venus, Machine Learning.

[24] Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A. (1997). Discovering Data Mining. Prentice-Hall.

[25] Caiado, J., Crato, N. Pe na, D. (2005). A periodogram-based metric for time series classification. Computational Statistics & Data Analysis.

[26] Costa Branco, P.J., Dente, J.A. (2001). Fuzzy systems modeling in practice. Fuzzy Sets and Systems 121, 73-93.

[27] Cotter, A. J. R., Buckland, S. T. (2004). Using the EM algorithm to weight data sets of unknown precision when modelling fish stocks. Mathematical Biosciences, 190, 1-7.

[28] Craven, M. W., Shavlik, J. W. (1997). Using neural networks for data mining. Future Generation Computer Systems, 13, 211-229.

[29] Delgado, A., Nirschl, H., Becker, T. (1996). First use of cognitive algorithms in investigations under compensated gravity. Microgravity Sci. and Techn. IX/3, 185-192.

[30] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B, 39, 1-38.

[31] Dhar, V., Stein, R. (1997). Seven Methods for Transforming Corporate Data into Business Intelligence. Prentice-Hall.

[32] Dubois, D., Prade, H. (1980). Fuzzy Sets and Systems - Theory and Applications. Academic Press.

[33] Esser, K. D. (1972). Zur Messung der Filtrierbarkeit. Monatsschrift für Brauerei 25, Nr. 6, S. 145-151.

[34] Esser, K. D. (1989). Probleme der Bierfiltration. Brautechnik aktuell. S. 177-180.

[35] Evers, H. (1996). Ersatz der Kieselgur bei der Bierfiltration durch ein regenerierbares Filterhilfsmittel. Berlin, Technische Universität, Dissertation.

[36] Fayyad, U. M, Piatetsky-Shapiro, G., Smyth, P. (1996). From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining S.1- 34.

[37] Fayyad, U.M, Piatetsky-Shapiro, G., Smyth, P. (1996). Knowledge Discovery and data mining: towards a unifying framework.Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 82-88.

[38] Fayyad, U.M., Smyth, P., Weir, N., Djorgovski, S. Automated analysis and exploration of image databases: Results, progress, and challenges. J. Intelligent Info. Syst. 4 (1995), 1-19.

[39] Finn, P., Muggleton, S., Page, D., Srinivasan, A. (1998). Pharamacophore discovery using the inductive logic programming system PROGOL. Maschine Learning.

[40] Gehring, H.; Bortfeldt, A. (2002). A parallel genetic algorithm for solving the container loading problem. International Transactions of Operational Research, Vol. 9 (2002) 4, S. 497-511.

[41] Geman, S., Bienenstock, E., Doursat, R. (1996). Neural networks and the bias/variabce dilema. Neural Computation 4, 1-58.

[42] Ghiaus, C. (2001). Fuzzy model and control of a fan-coil. Energy and Buildings 33, 545-551.

[43] Glass, K., Karopka, T. (2002). Genomic Data Explosion The Challenge for Bioinformatics? Advances in Data Mining, S.80-98.

[44] Grabmeier, J., Rudolph, A. (2001). Techniques of Cluster Algorithms in Data Mining. Data Mining and Knowledge Discovery.

[45] Groth, R. (1998). Data Mining: A Hands-On Approach for Business Professionals. Prentice-Hall.

[46] Jolliffe, I. T. (1986). Principal component analysis. Springer-Verlag, Berlin.

[47] Hackel, A., Heidenreich, E., Hölinger, W., Tittel, R. Filterhilfsmittelfiltration. VDI-Verlag, Fortschr.-Berichte, Reihe 3, Nr. 348.

[48] Hand, D. J. (2002). Statistics. Handbook of Data Mining and Knowledge Discovery. Oxford University Press.

[49] Hansen, J. V., Nelson, R. D. (2000). Data mining of time series using stacked generalizers. Neurocomputing.

[50] Hartig, J. A., Wong, M. A. (1979). A k-means clustering algorithm. Applied Statistics 28, 100-108.

[51] Henrion, R., Henrion, G. (1995). Multivariate Datenanalyse. Springer-Verlag Berlin.

[52] Hipp, J., Gntzer, U. und Nakhaeizadeh, G. (2002). Data Mining of Association Rules and the Process of Knowledge Discovery in Databases. Advances in Data Mining. ISBN 3-540-44116-6, S.15-35.

[53] http://www.kdnuggets.com

[54] Huber, E. (1997). Die Trübungsmessung: Grundlagen, Anwendung und Grenzen. Sonderdruck der Sigrist-Photometer AG, Ennetbürgen.

[55] Kafka, C. (1999). Konzeption und Umsetzung eines Leitfadens zum industriellen Einsatz von Data-Mining. Wbk, Forschungsberichte aus dem Institut für Werkzeugmaschinen und Betriebstechnik der Universit Karlsruhe. Band 99.

[56] Kaufmann, L., Rousseeuw, p. J. (1990). Finding groups in data: an introduction to cluster analysis. Wiley series in probability ans mathematical statistics.

[57] Kilimann, K. V., Hartmann, C., Delgado, A., Gänzle, M. G. (2004). Data mining and fuzzy modelling of high pressure inactivation pathways of Lactococcus lactis. 3rd International Conferences on High Pressure Bioscience and Biotechnology, Rio de Janeiro, Brazil.

[58] Kilimann, K. V., Hartmann, C., Delgado, A., Vogel, R. F., Gänzle, M. G. (2005). A Fuzzy logic based model for the multi-stage high-pressure inactivation of Lactococcus lactis ssp. Cremoris MG 1363, International Journal of Food Microbiology 98 (1), pp. 89-108.

[59] Kilimann, K. V., Hartmann, C., Gänzle, M. G.(2004). Application of fuzzy logic and statistical analysis for modelling high-pressure induced microbial inactivation. (P. Neittaanm¨aki, T. Rossi, K. Majava, O. Pironneau) 4th European Congress on Computational Methods in Applied Sciences and Engineering (Eccomas 2004), Jyväskylä/Finnland.

[60] Kim, W., Choi, B.-J., Hong, E.-K., Kim, S.-K., Lee, D. (2001). A Taxonomy of Dirty Data. Data Mining and Knowledge Discovery.

[61] Klösgen, W. (2002). Overview of discovery systems. Handbook of Data Mining and Knowledge Discovery. Oxford University Press.

[62] Klösgen, W., Żytkow, J. M. (2002). Handbook of Data Mining and Knowledge Discovery. Oxford University Press.

[63] Kohavi, R., Wolpert, D. H. (1996). Bias plus variance decomposition for zero-one loss functions. Machine Learning: Proceedings of the 13th International Conference, 275-283.

[64] Kononenko, I., Hong, S. J. (1997). Attribute selection for modelling. Future Generation Computer Systems, 13, 181-195.

[65] Kreisz, S. (2003). Der Einfluss von Polysacchariden aus Malz, Hefe und Bakterien auf die Filtrierbarkeit von Würze und Bier. München, Technische Universität, Dissertation.

[66] Kreisz, S. (2004). Oral communication.

[67] Krone, A. (1999). Datenbasierte Generierung von relevanten Fuzzy-Regeln zur Modellierung von Prozesszusammenhängen und Bedienstrategien. VDI-Verlag GmbH, Düsseldorf.

[68] Kubat, M., Holte, R. C., Matwin, S. (1998). Machine Learning for the detection of oil spills in satellite radar images. Maschine Learning.

[69] Kurz, T., Arnold, S., Fellner, M., Murnleitner, E., Mitzscherling, M., Becker, T., Delgado, A. (2002). Modelling and Optimisation Approaches in Industrial Fermentation Prosesses. Proceedings of the 2nd International Conference on Simulation in Food and Bio Industries, pp. 86-90.

[70] Kurz, T., Mieleitner, J., Becker, T., Delgado, A. (2002). Model based simulation of the brewing yeast propagation , Journal of the Institute of Brewing 108 (2), 248-255.

[71] Kusche, M. (2005). Kolloidale Trübung in untergärigen Bieren. München, Technische Universität, Dissertation.

[72] Langley, P. and Simon, H. (1995). Applications of Machine Learning an Rule Induction. Communcications of th ACM. Nov 1995/ Vol. 38, No.11.

[73] Lee, Y., Buchanan, B. G., Aornis, J: M: (1998). Knowledge-based learning in exploratory science: Learning rules to predict rodent carcinogenicity. Maschine Learning.

[74] Leech, W.J. A rule-based process control method with feedback. Advances in Instrumentation 42 (1986), 169-175.

[75] Lindemann, B., Fontaine, J. und Krüger, E. (1991). Der Einfluss filtrationshemmender Stoffe auf die Vorhersage der Filtrierbarkeit des Bieres. Teil 1: Einflussfaktoren auf das Ergebnis des Membranfiltertests nach Esser. Monatsschrift für Brauwissenschaft. Nr. 44, S. 336-340.

[76] Liu, Z., Almhana, J., Choulakian, V., McGorman, R. (2006). Online EM algorithm for mixture with application to internet traffic modeling. Computational Statistics & Data Analysis, 50, 1052-1071.

[77] Lloyd-Williams, M. (1999). Empirical studies of the knowledge discovery approach to health-information analysis. The Institute of Electrical Engineers. Professional Applications of Computing Series 1: Knowledge Discovery and Data Mining. S.139-158.

[78] Lozano, M. A., Escolano, F. (2005). Protein classification by matching and clustering surface graphs. Pattern Recognition.

[79] Martens, H., Næs, T. (1989). Multivariate Calibration. Wiley. New York.

[80] McLachlan, G. J., Krishnan, T. (1997). The EM Algorithm and Extensions. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

[81] Menzies, T., Hu, Y. (2003). Data Mining For Very Busy People. IEEE Computer, 18-25.

[82] Michaud, P. (1996). Clustering techniques. Future Generation Computer Systems.

[83] Michie, D.(1989). Problems of computer-aided concept formation. J.R.Quinlan, Ed., Apllications of Expert Systems. Addison-Wesley, Wokingham, UK.

[84] Murnleitner, E. (2002). State Detection and Feedback Control of the Anaerobic Wastewater Treatment using Fuzzy Logic. Lehrstuhl für Fluidmechanik und Prozessautomation, TU München, Dissertation.

[85] Nakhaeizadeh, G. (2000). Vom Data Mining zum verteilten Data Mining. Computional Intelligence im industriellen Einsatz. VDI Berichte Nr. 1526 S.199-209.

[86] Narziß, L. (1995). Abrißder Bierbrauerei. Ferdinand Enke Verlag, Stuttgart.

[87] Narziß L. und Heiden, L. (1971). Über die Anwendung von Sauermalz oder Sauerwürze beim Sudprozess. Brauwelt 111, Nr. 22, S. 435-440.

[88] NarzißL. und Heiden, L. (1972). Über die Anwendung von Sauermalz beim Sudprozess. 2. Mitteilung, Zusatz verschieden hoher Sauermalzgaben bei unterschiedlicher Härte des Brauwassers. Brauwelt 112, Nr. 18, S. 335-340.

[89] Pölmann, R. (1977). Zum Stand der Kenntnis ber die Filtrierbarkeit der Biere. Brauerei Journal Nr. 13 /1977. S.330-333.

[90] Papp, A. (2000). Erarbeitung von Beurteilungskriterien zur Vorhersage der Stabilisierbarkeit und der chemisch-physikalischen Stabilität im unfiltrierten Bier. München, Technische Universität, Dissertation.

[91] Provost, F., Kohavi, R. (1998). On applied research in Machine Learning. Machine Learning, vv, 1-6.

[92] Quinlan, J. R. (1986). Induction of decision trees, Machine Learning, 1, 1986.

[93] Raible, K. (1990). Eine einfache neue Methode zur Bewertung der Filtrationseigenschaften von Bier. Brauwissenschaft 43, Nr. 2, 60-65.

[94] Raible, K., Bantleon, H. (1968). Über die Filtrationseigenschaften von Bier. Mschr. Brauerei 21, Nr. 10, 277-285.

[95] Raible, K., Heinrich, Th., und Niemsch, K. (1990). Eine einfache, neue Methode zur Bewertung der Filtrationseigenschaften von Bier. Monatsschrift für Brauwissenschaft 43/1990. Nr.2, S.60-65.

[96] Rajan, J. J., Rayner, P. J. W. (1995). Unsupervised time series classification. Signal Processing.

[97] Reed, R. (1983). Beer filterability Evaluation and alternative approaches to processing. Proc. of the 19th EBC-Congress, London, 225-232.

[98] Reinartz, T. (2002). Stages of the Discovery Process. Handbook of Data Mining and Knowledge Discovery.

[99] Rinne, H. (2003). Taschenbuch der Statistik. Verlag Harry Deutsch.

[100] Sang, K. (2002). Overview of SAS Enterprise Miner. Handbook of Data Mining and Knowledge Discovery. Oxford University Press.

[101] Schmidt, M., Eder, C., Delgado, A. (2005). Automatische Schadenserkennung an Getränkekästen mittels optischer Schwingungsmessung und Neuronumerik. 13. Fachtagung "Lasermethoden in der Strömungsmesstechnik" GALA 2005, Cottbus.

[102] Schmidt, N. A. (2002). Verbesserung der filtrationstechnischen Eigenschaften von Filterhilfsmitteln durch ein thermisches Verfahren. München, Technische Universität, Dissertation.

[103] Schur, F. (1978). Charakterisierung filtrationshemender Stoffe. Brauerei Rundschau. Jg. 89, Nr. 8.

[104] Shearer, C., Caron, P. (2002). Clementine. Handbook of Data Mining and Knowledge Discovery. Oxford University Press.

[105] Siebert, K. J. (1984). Assessment of filterability of beer and wort. MBAA Technical Quaterly 21, Nr. 3, 112-123.

[106] Sii, H.S., Ruxton, T. and Wang, J. (2001). A fuzzy-logic-based approach to qualitative safety modelling for marine systems. Reliability Engineering and System Safety 73, 19-34.

[107] Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., Hjalmarssom, H., Juditsky, A. (1995). Nonlinear Black-box Modeling in References 70 Systems Identification: a Unified Overview. Automatica, vol. 31, no. 12, 1691-1724.

[108] Struyf, A., Hubert, M., Rousseeuw, P. J. (1997). Integrating Robust Clustering Techniques in S-PLUS. Computational Statistics and Data Analysis 26, 17-37.

[109] Sugeno, M, Yasukawa, T. (1993). A Fuzzy-Logic-Based Approach to Qualitative Modeling. IEEE Transactions on Fuzzy Systems 1(1), 7-31.

[110] Thausing, J. (1877). Malzbereitung und Bierfabrikation. Leibzig.

[111] The R Project for Statistical Computing, http://www.r-project.org/.

[112] Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., De Yong, S., Lewi, P., Smeyers-Verbeke, J. (1998). Data handling in science and technology. Handbook of chemometrics and qualimetrics Part A/Part B. Elsevier Science B.V.

[113] Viehhauser, W. (2005). Oral communication.

[114] Warren Liao, T. (2005). Clustering of time series data — a survey. Pattern Recognition.

[115] Watanabe, M. Yamaguchi, K. (2004). The EM Algorithm and Related Statistical Models. Macel Dekker, Inc.

[116] Webster, R. D. J., Malzahn, S. W. (1983). The influence of beer quality on filtration performance. Proc. of the 19th EBC-Congress, London, 233-240.

[117] Weiss, S. M., Zhang, T. (2003). Performace Analysis and Evaluation. Handbook of Data Mining. Lawrence Erlbaum Associates.

[118] Weiss, S., Indurkhya, N. (1998). Predictive Data Mining. Morgan Kaufmann.

[119] Witten, I. H. und Frank, E. (2001). Data Mining. Praktische Werkzeuge und Techniken für das maschinelle Lernen. Carl Hanser Verlag. ISBN 3-446-21533-6.

[120] Xiong, Y., Yeung, D.-Y. (2003). Time series clustering with ARMA mixtures. Pattern Recognition.

[121] Yager, R.R., Filev, D.P. (1994). Essentials of Fuzzy Modeling and Control. John Wiley & Sons Inc., New York.

[122] Ye, N. (2003). Handbook of Data Mining. Lawrence Erlbaum Associates.

[123] Zacharias, J., Delgado, A. (2003). Online Detektion defekter Gebinde mittels Neuronumerik. 36. Technologisches Seminar Weihenstephan.

[124] Zacharias, J., Hartmann, Ch., Delgado, A. (2001). Recognition of Damages on Crates of Beverages by an Artificial Neural Network. Accepted in Proceedings of EUNITE 2001 (European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems).

[125] Zadeh, L. A. (1965). Fuzzy Sets. Information and Control.

[126] Zadeh, L. A. (1968). Fuzzy algorithms. Information and Control, 12, 94-102.

[127] Zadeh, L. A. (1973). Outline of A New Approach to the Analysis of Complex Systems and Decision Processes.

[128] Zadeh, L.A. (1996). The roles of fuzzy logic and soft computing in the conception, design and deployment of intelligent systems. BT Technol J. 14, No. 4, 32-36.

[129] Zaït, M., Messatfa, H. (1996). A Comparative study of clustering methods. Future Generation Computer Systems.

[130] Zell, A. (1994). Simulation neuronaler Netze. Addison Wesley Longman Verlag GmbH.

[131] Zimmermann, H.-J., Angstenberger, J., Lieven, K., Weber, R. (1993) Fuzzy-Technologien: Prinzipien, Werkzeuge, Potentiale. VDI Verlag Düsseldorf.

Appendix A

Reference Plant

The Data used for this work is taken from an local brewery. The reference plant is a modern candle filter, which was implemented in 1999. The reference brewery produces different beers, from which eleven are filtered using the mentioned reference plant. During one batch, which takes normally one production day, up to six different beer types (mostly two or three) are processed. Three different kieselguhr types are used as filter aids. The filtration is followed by stabilisation and a particle filter, before it is bottled.

Chronological arranged in front of the filtration is the fermentation, conditioning and maturation. Due to the historical growth of the reference brewery, these production steps are carried out in different stock cellars, each with different tanks and different tank architectures. For the same reason, there are three brew houses.

The data can be divided in two parts: the data accumulating before filtration, like laboratory data and information of the previous process steps, and data generated during the filtration process.

Appendix B

Classification of Filtration - First Approach

B.1 Cluster Analysis

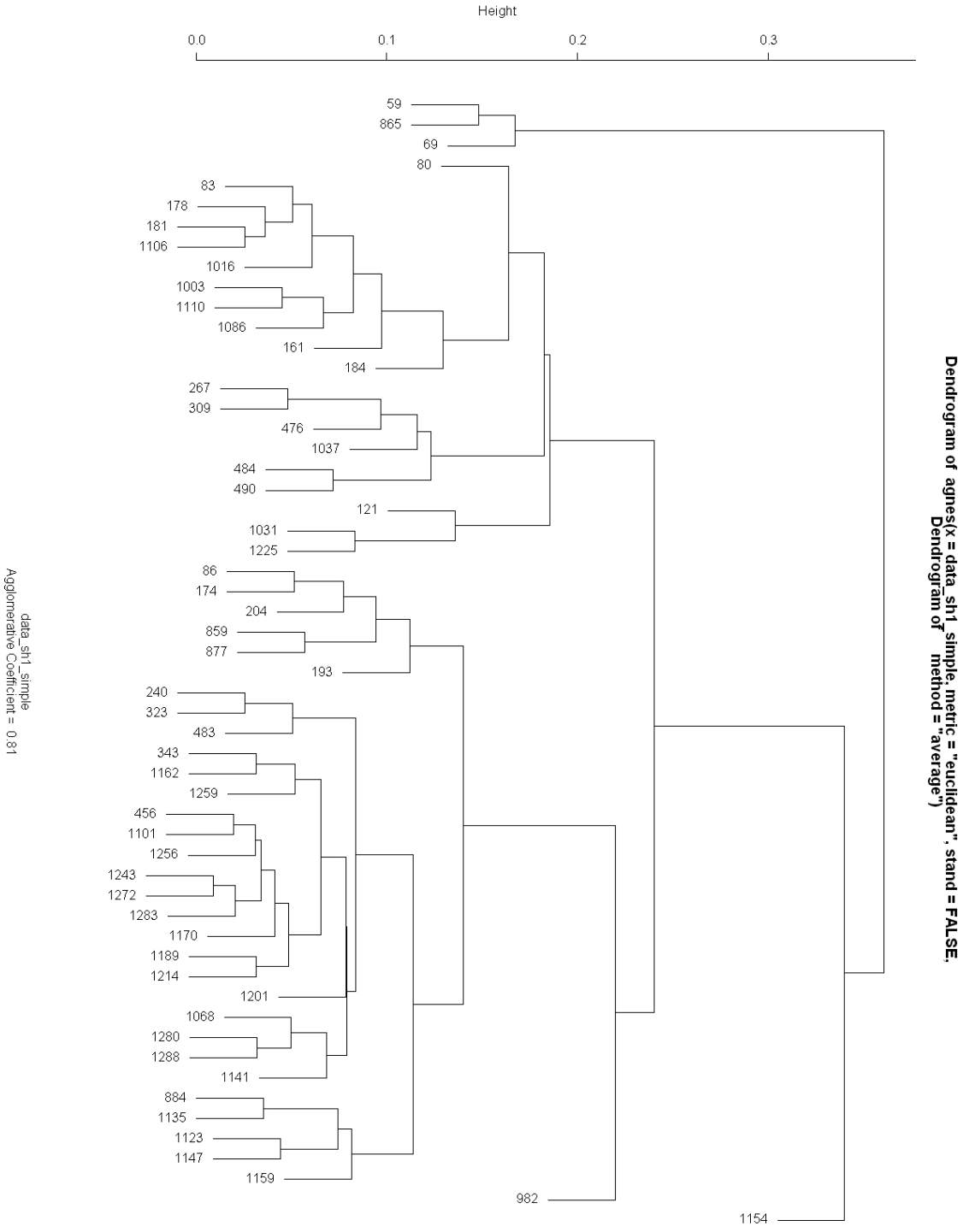


Figure B.1: Dendogram of Filtration Data using AGNES with method "average"

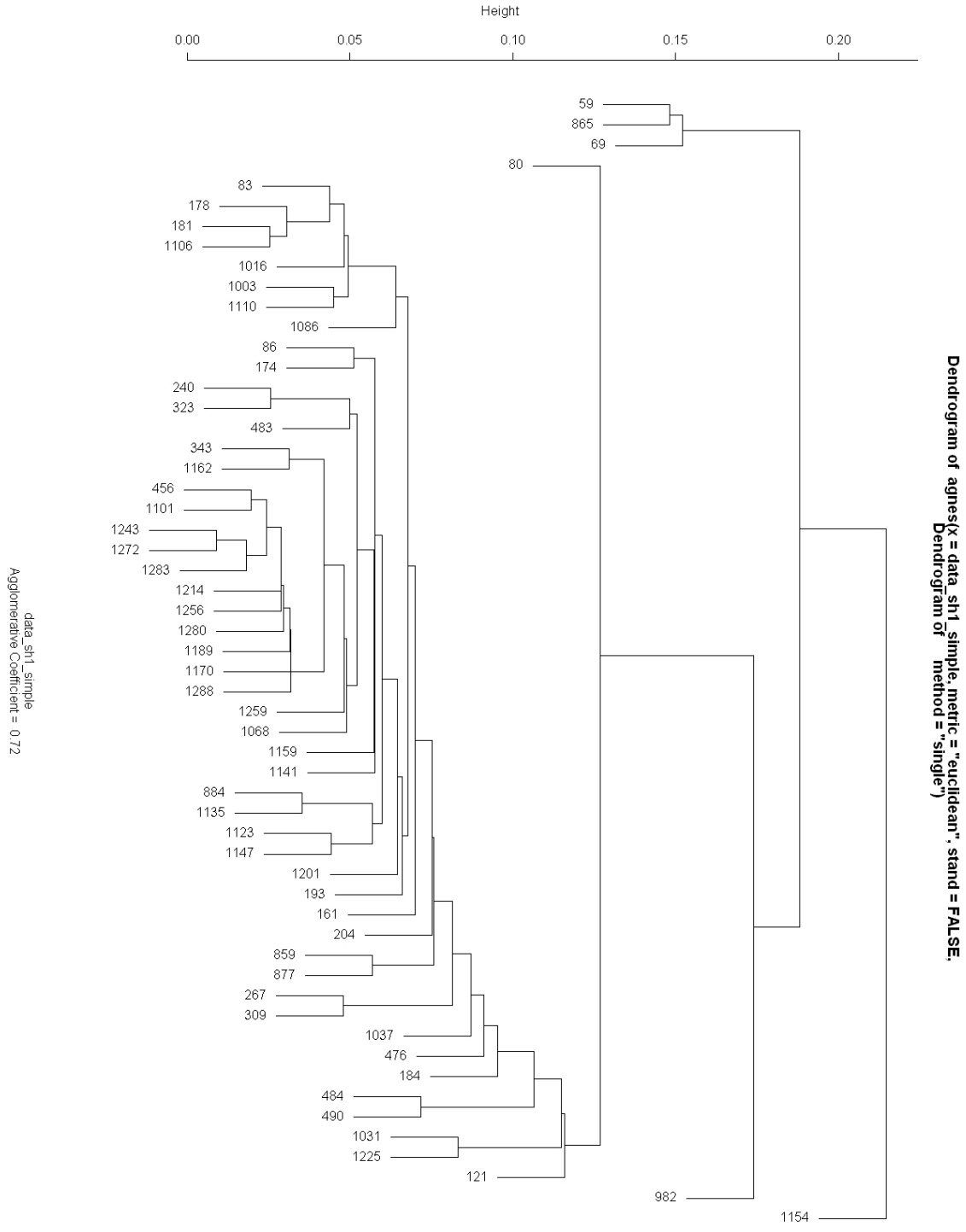


Figure B.2: Dendogram of Filtration Data using AGNES with method "single"

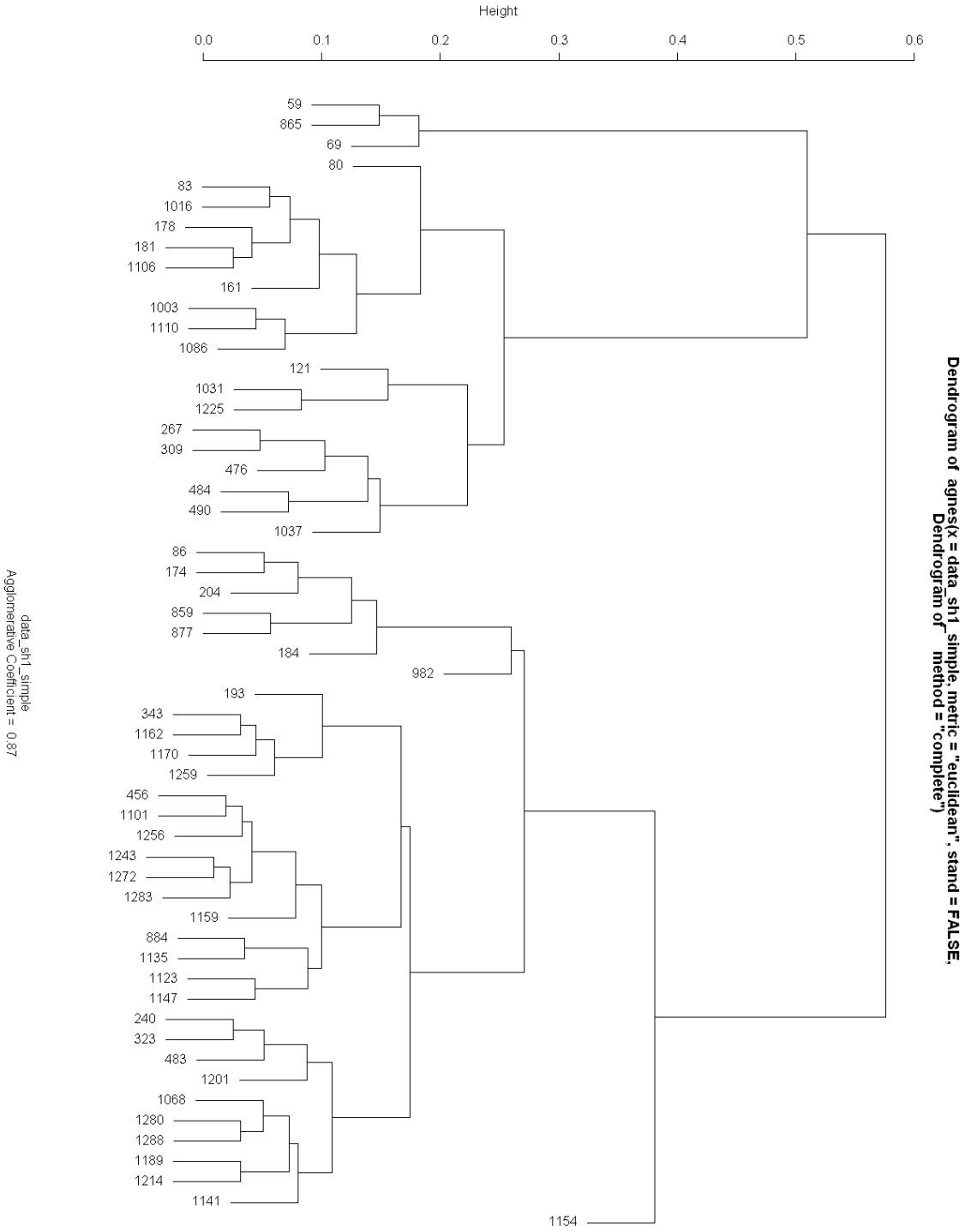


Figure B.3: Dendogram of Filtration Data using AGNES with method "complete"

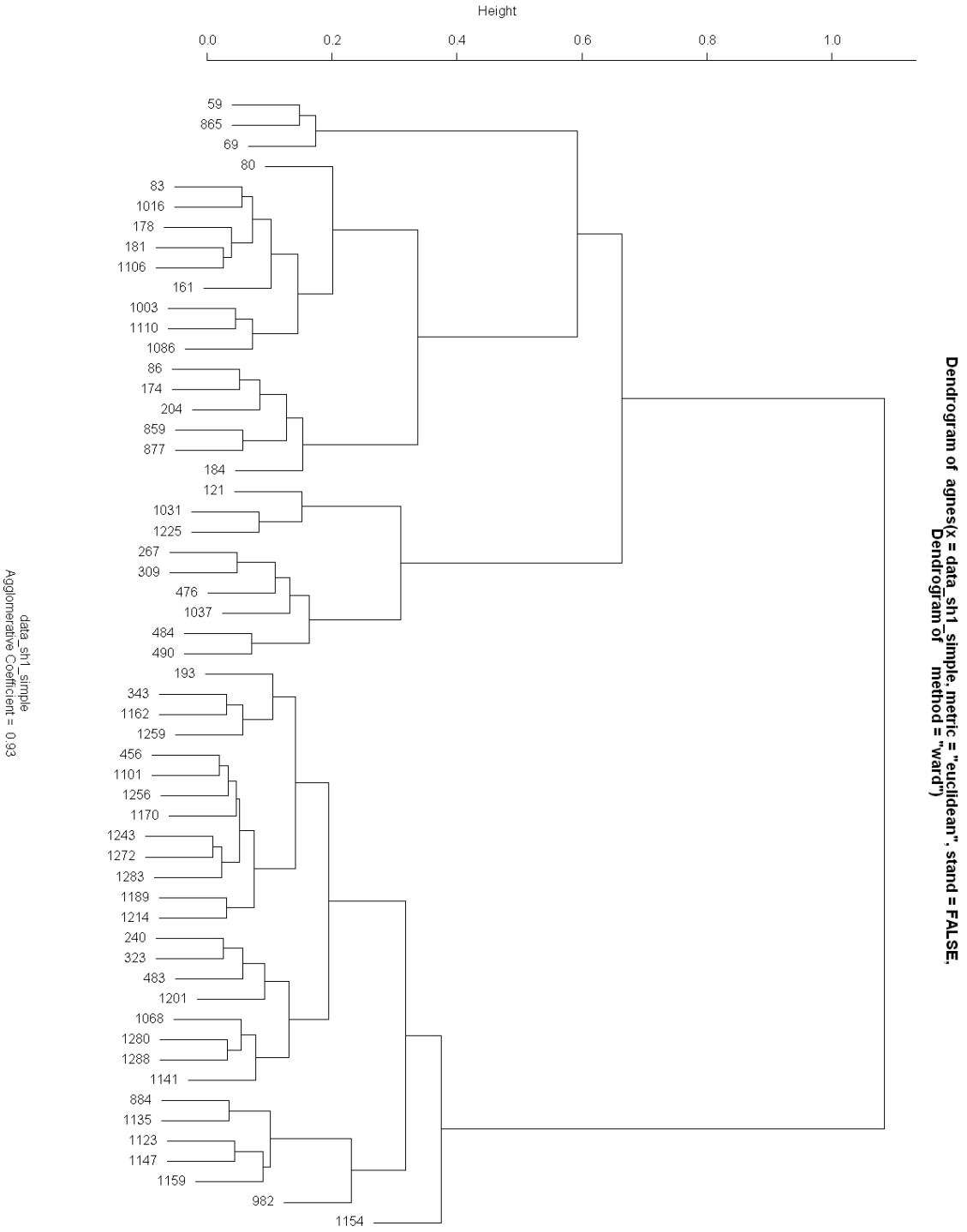


Figure B.4: Dendogram of Filtration Data using AGNES with method "ward"

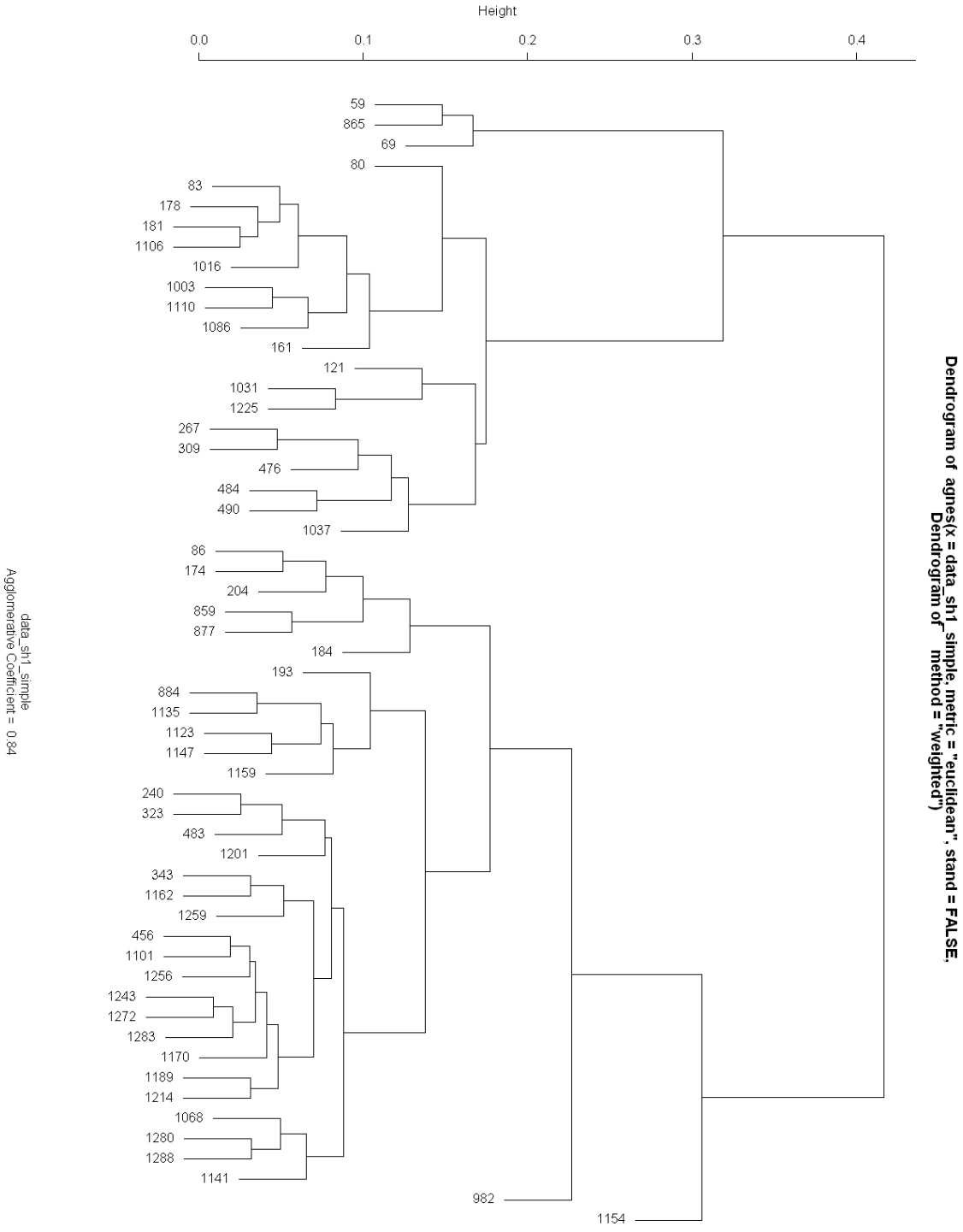


Figure B.5: Dendogram of Filtration Data using AGNES with method "weighted"

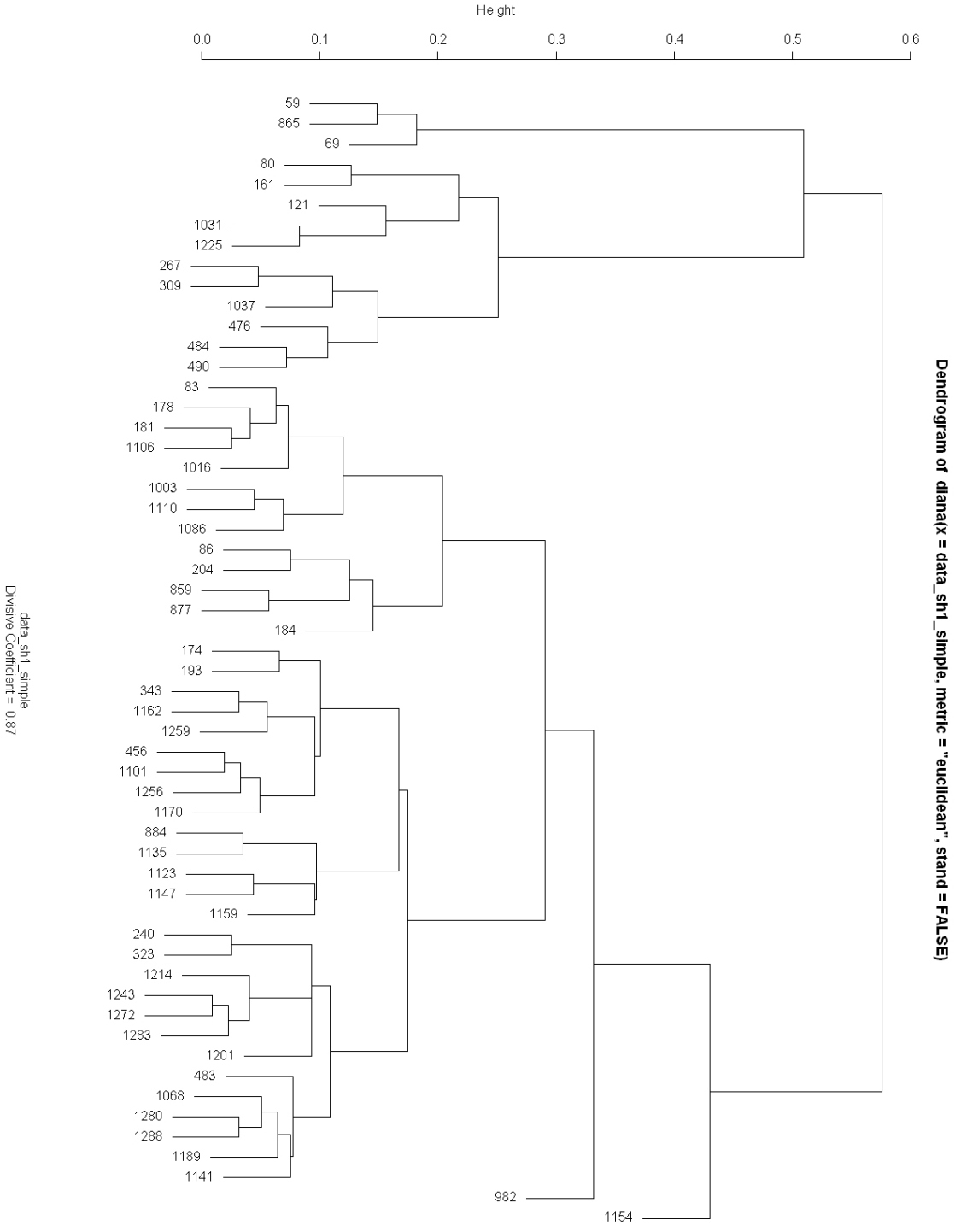


Figure B.6: Dendogram of Filtration Data using DIANA

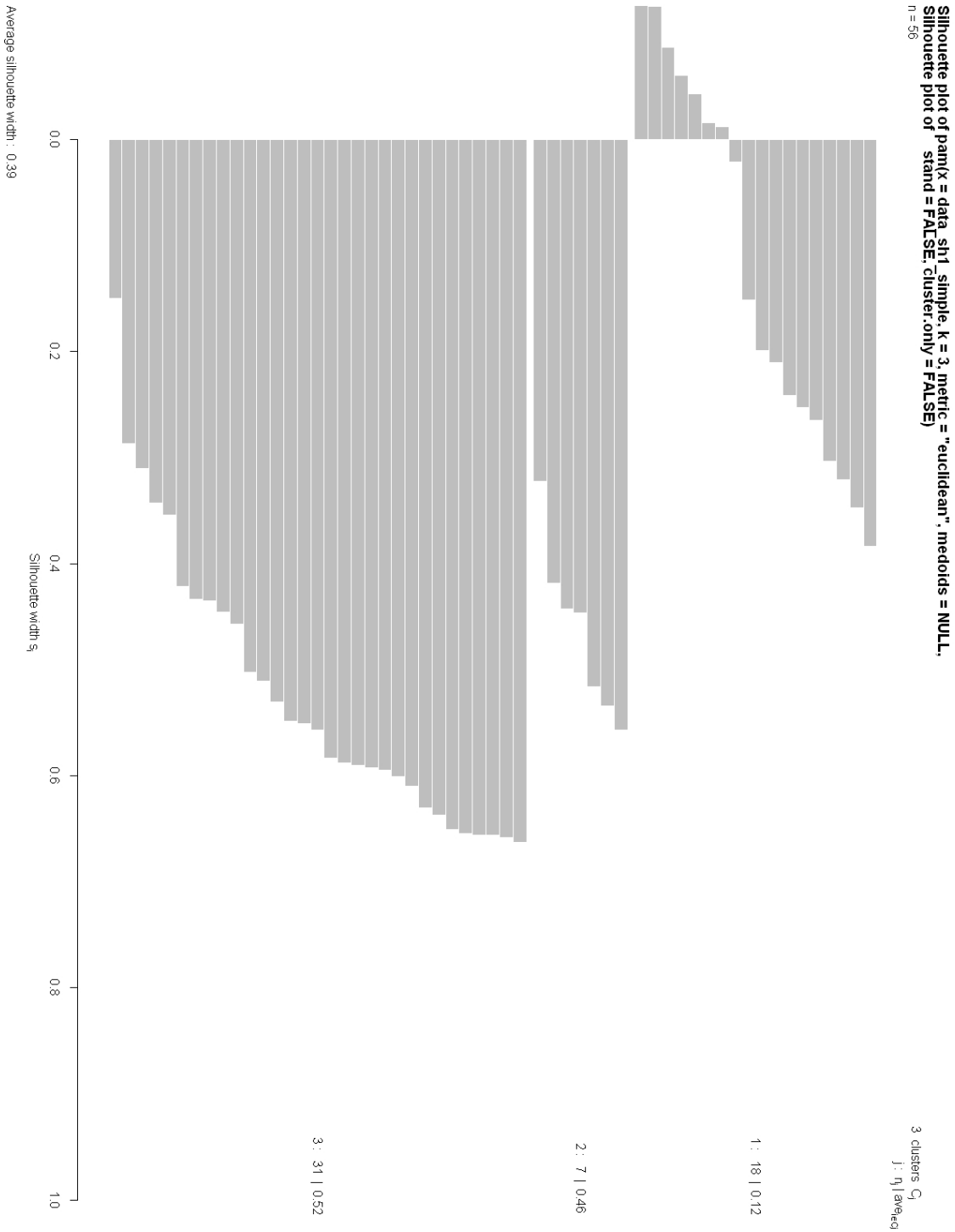


Figure B.7: Silhouette Plot of Filtration Data using PAM, $n = 3$ cluster

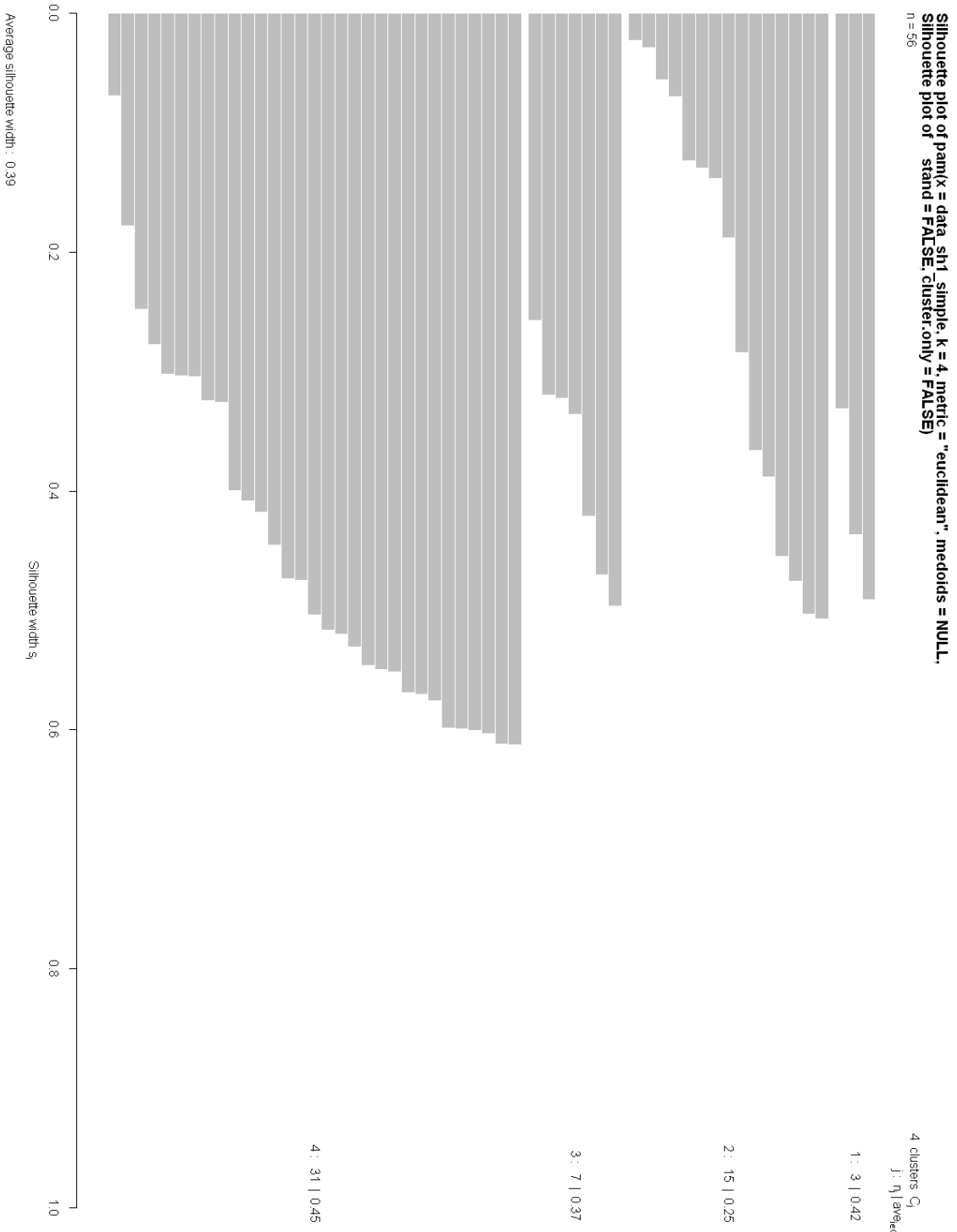


Figure B.8: Silhouette Plot of Filtration Data using PAM, $n = 4$ cluster

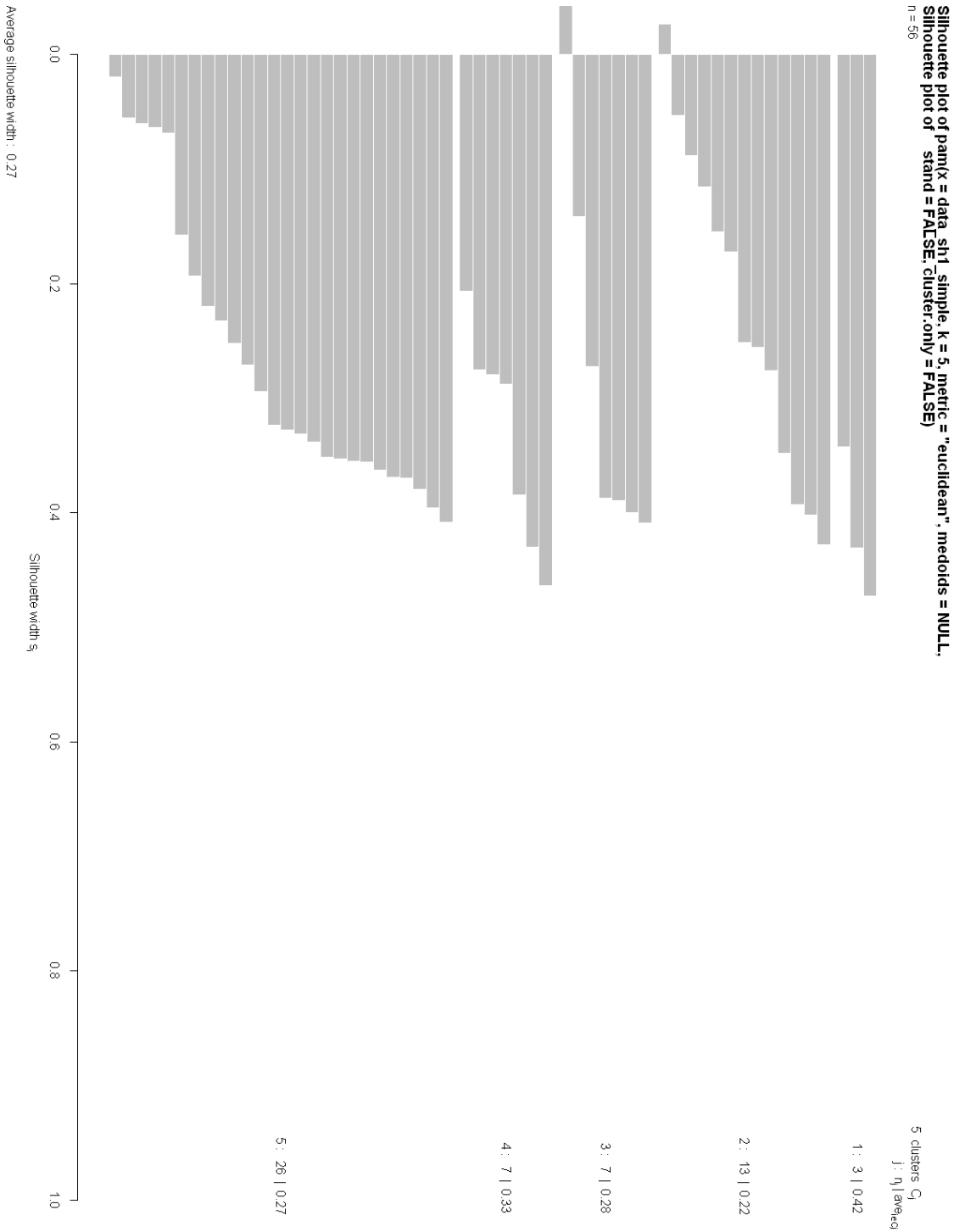


Figure B.9: Silhouette Plot of Filtration Data using PAM, $n = 5$ cluster

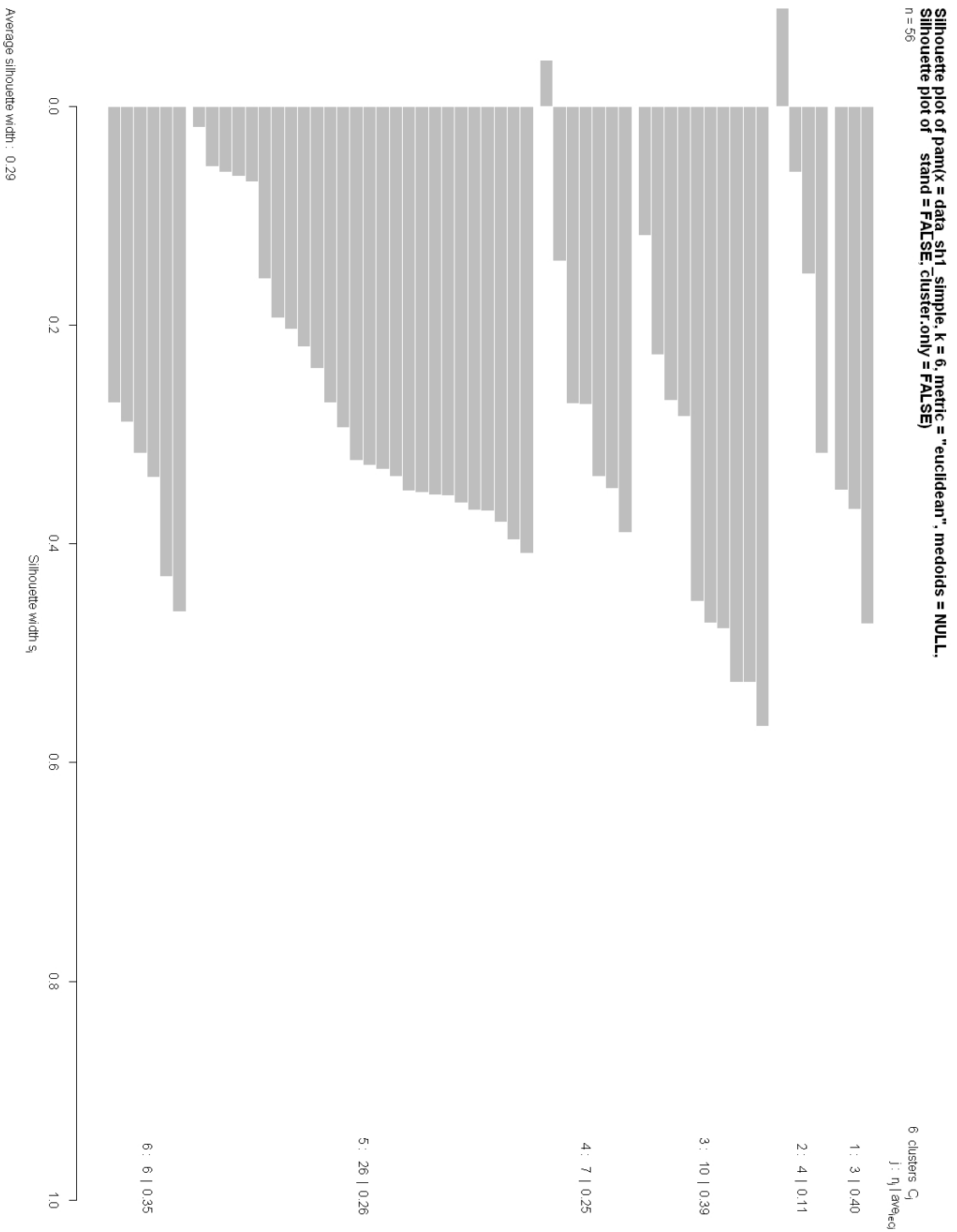


Figure B.10: Silhouette Plot of Filtration Data using PAM, $n = 6$ cluster

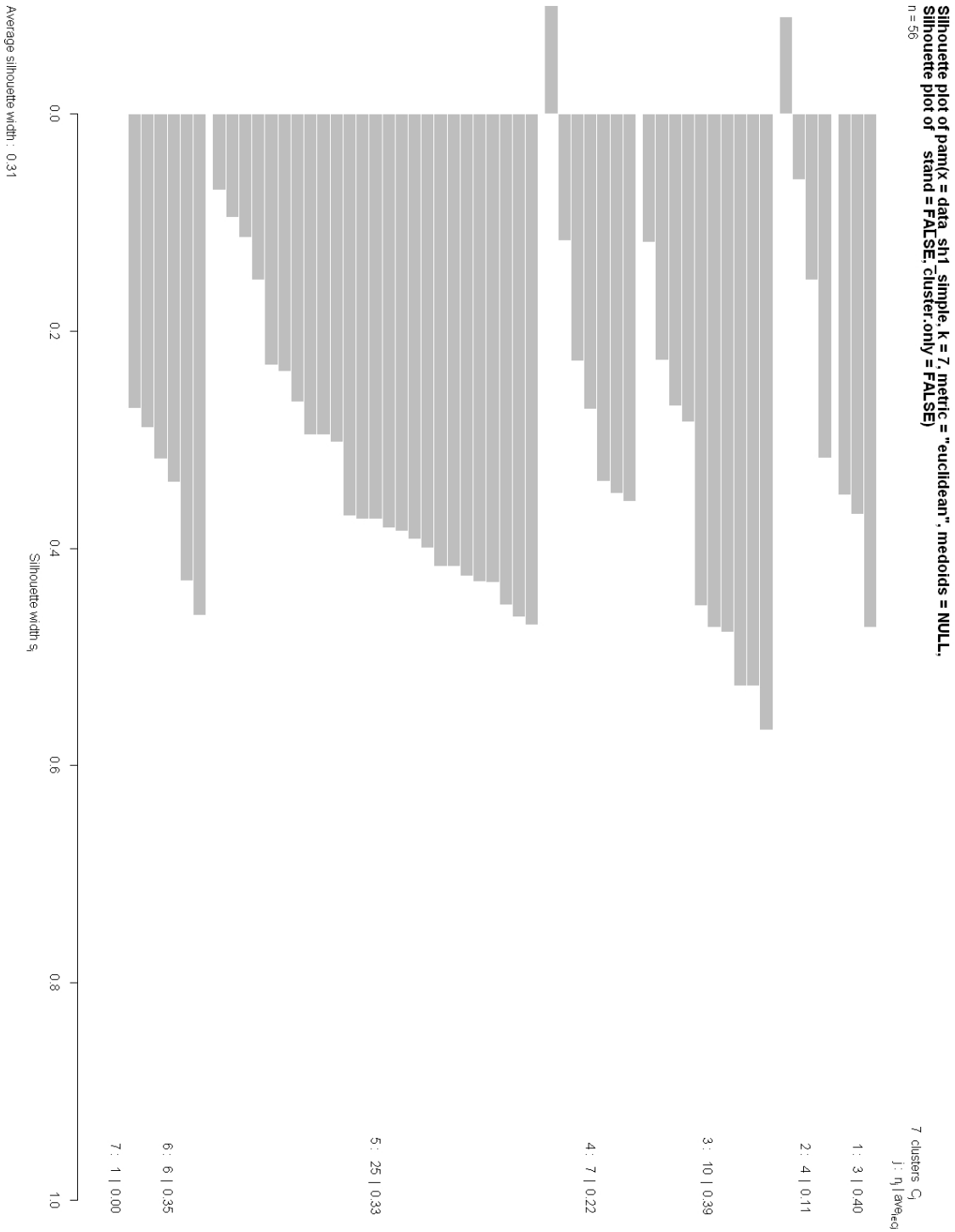


Figure B.11: Silhouette Plot of Filtration Data using PAM, $n = 7$ cluster

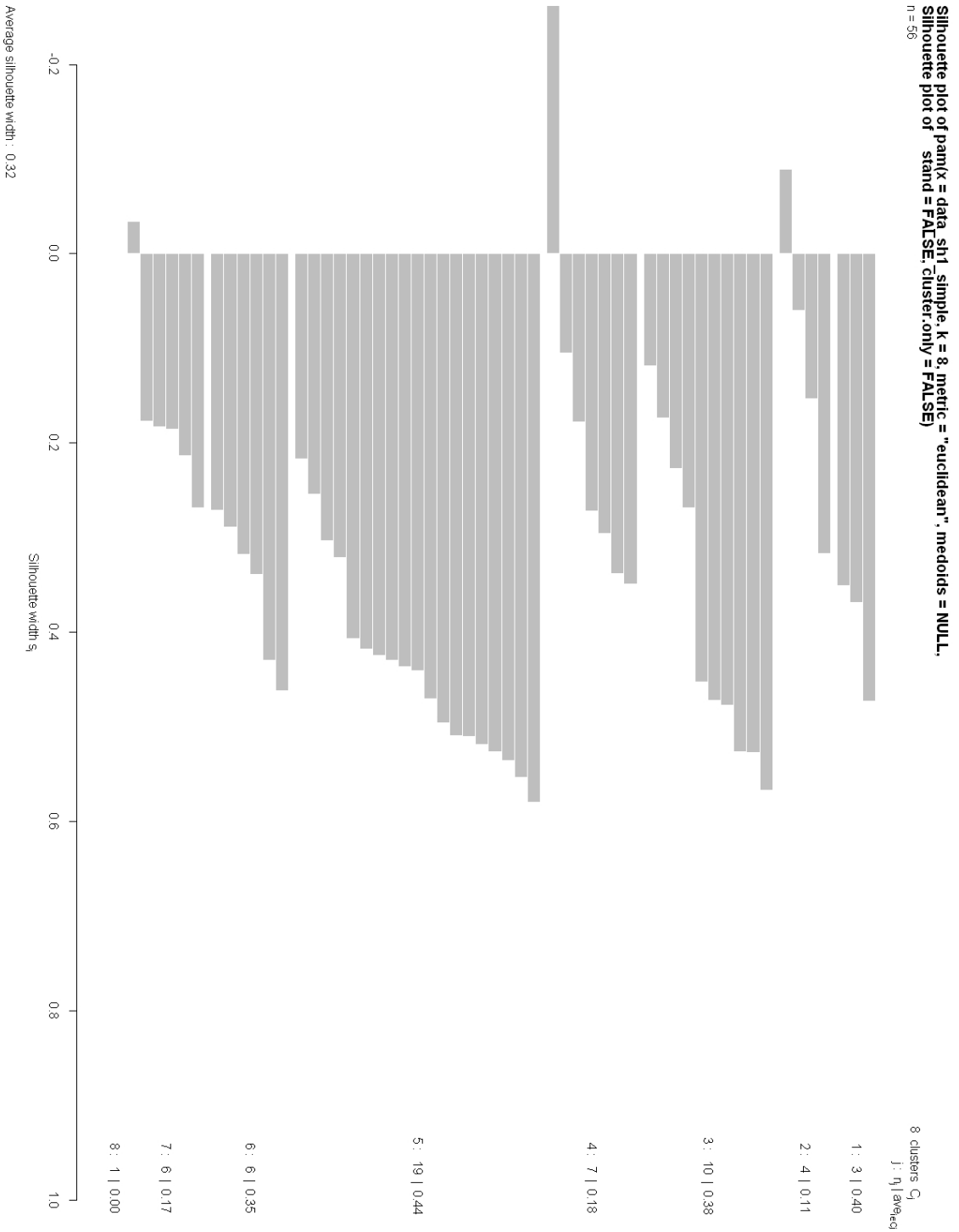


Figure B.12: Silhouette Plot of Filtration Data using PAM, $n = 8$ cluster

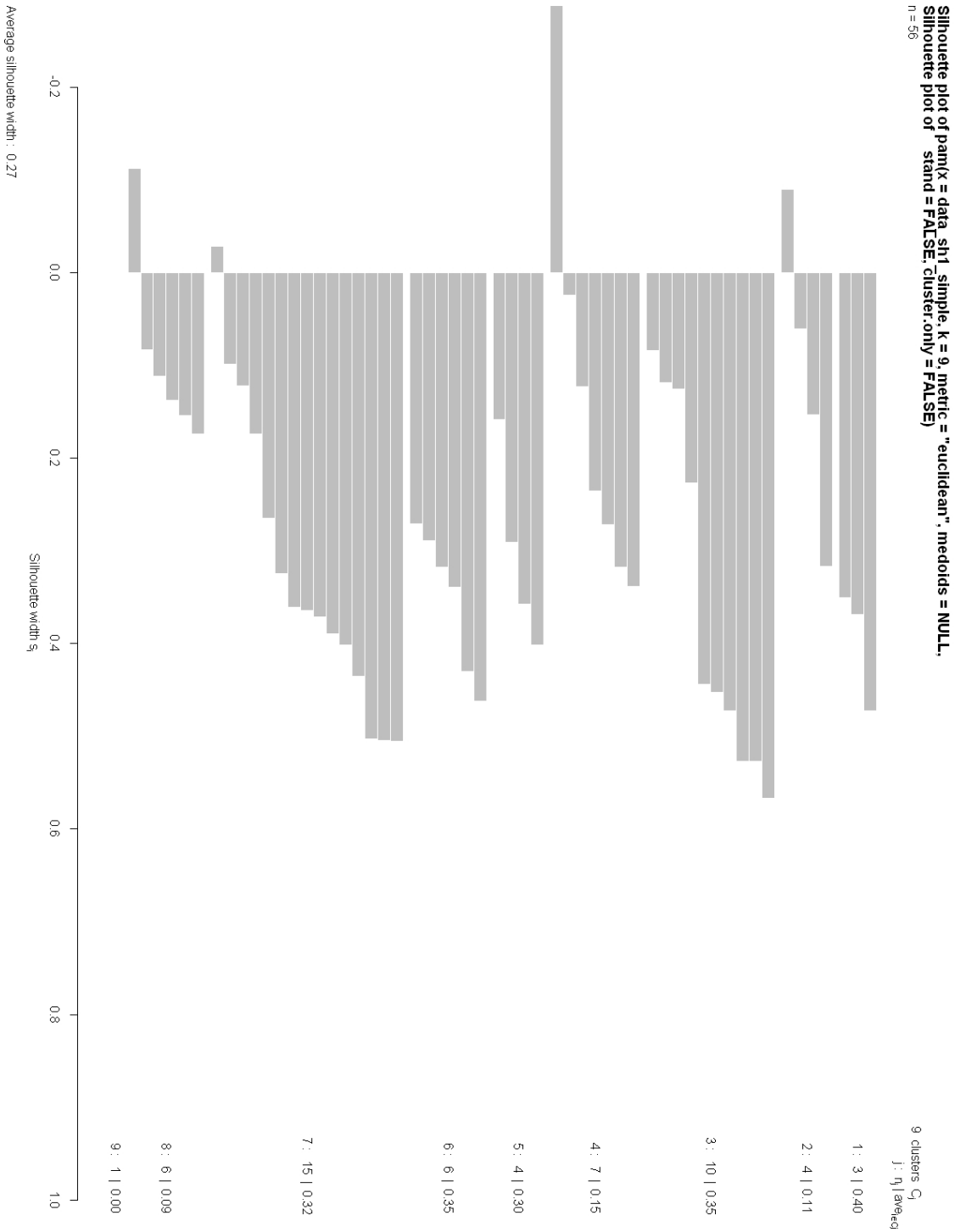


Figure B.13: Silhouette Plot of Filtration Data using PAM, $n = 9$ cluster

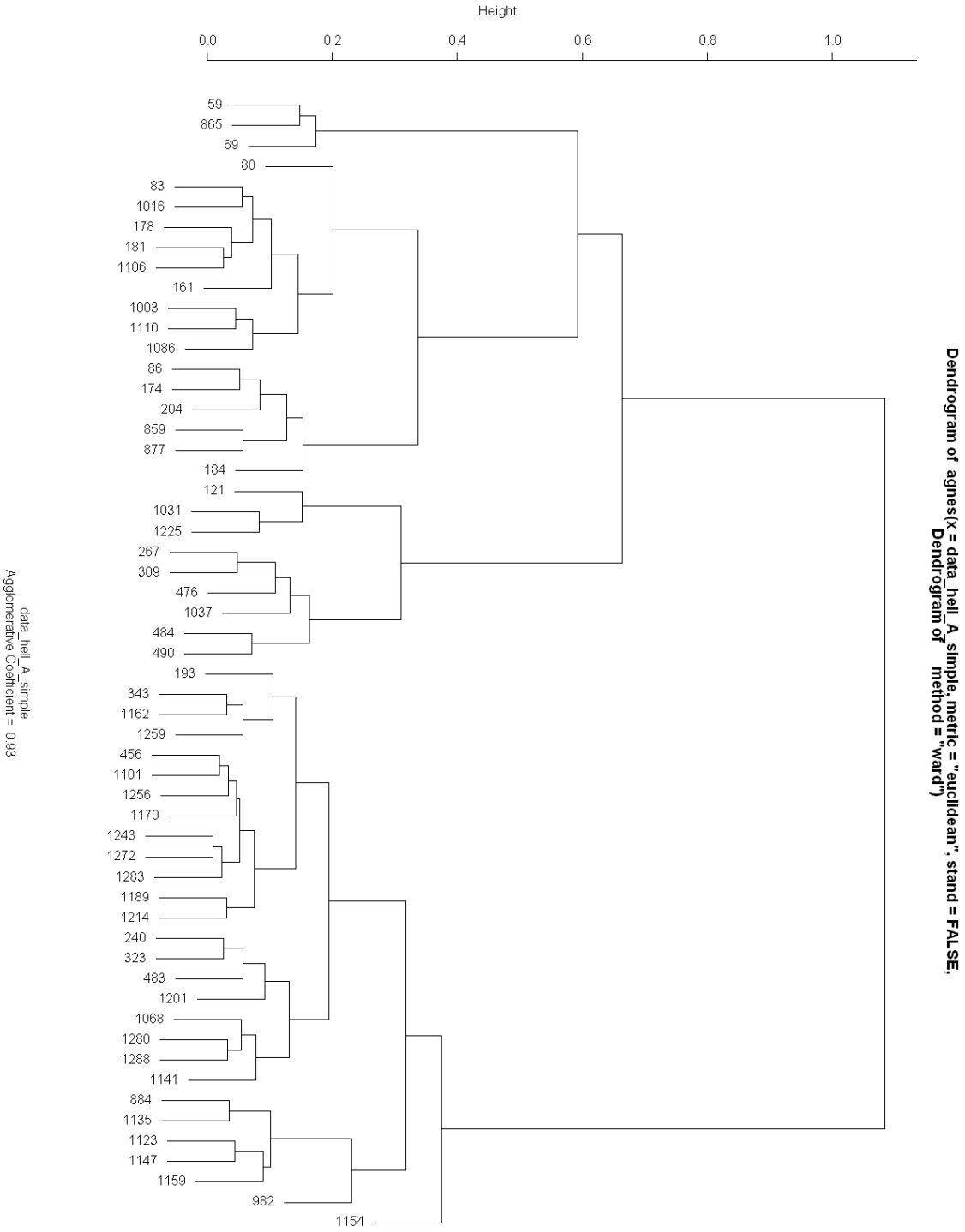


Figure B.14: Dendogram of Filtration Data from "Hell_A" using AGNES and Ward's Method

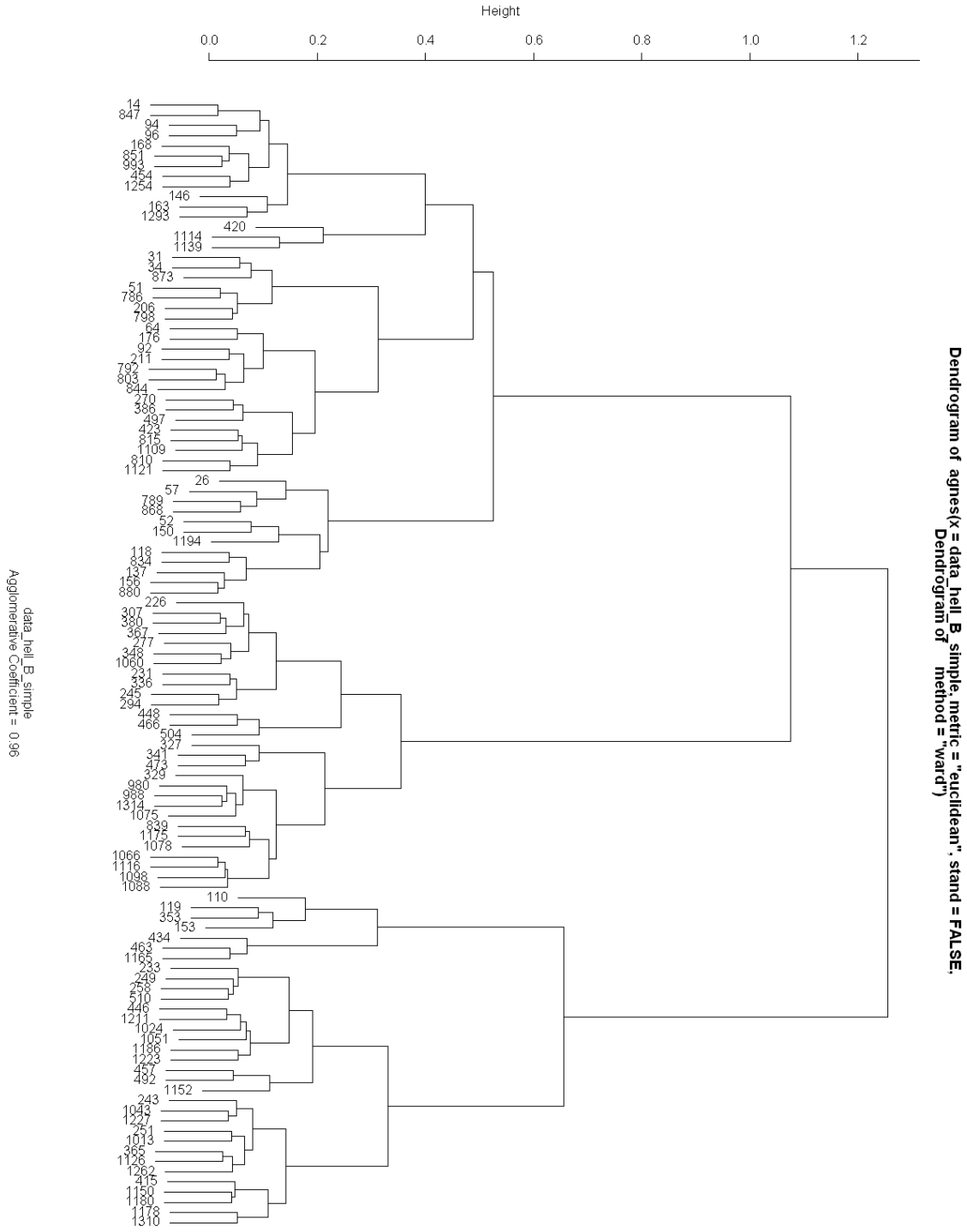


Figure B.15: Dendogram of Filtration Data from "Hell_B" using AGNES and Ward's Method

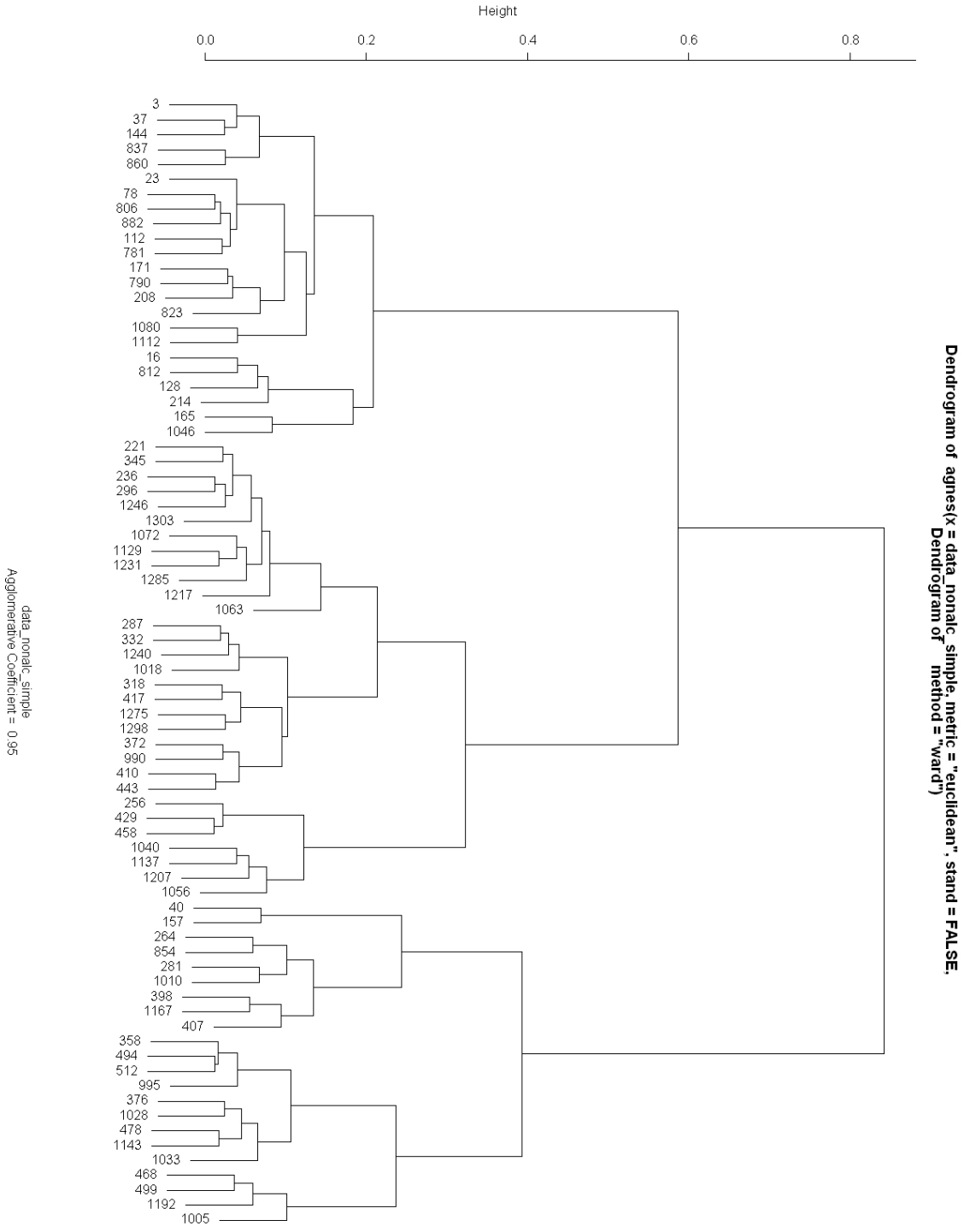


Figure B.16: Dendogram of Filtration Data from "Nonalcoholic" using AGNES and Ward's Method

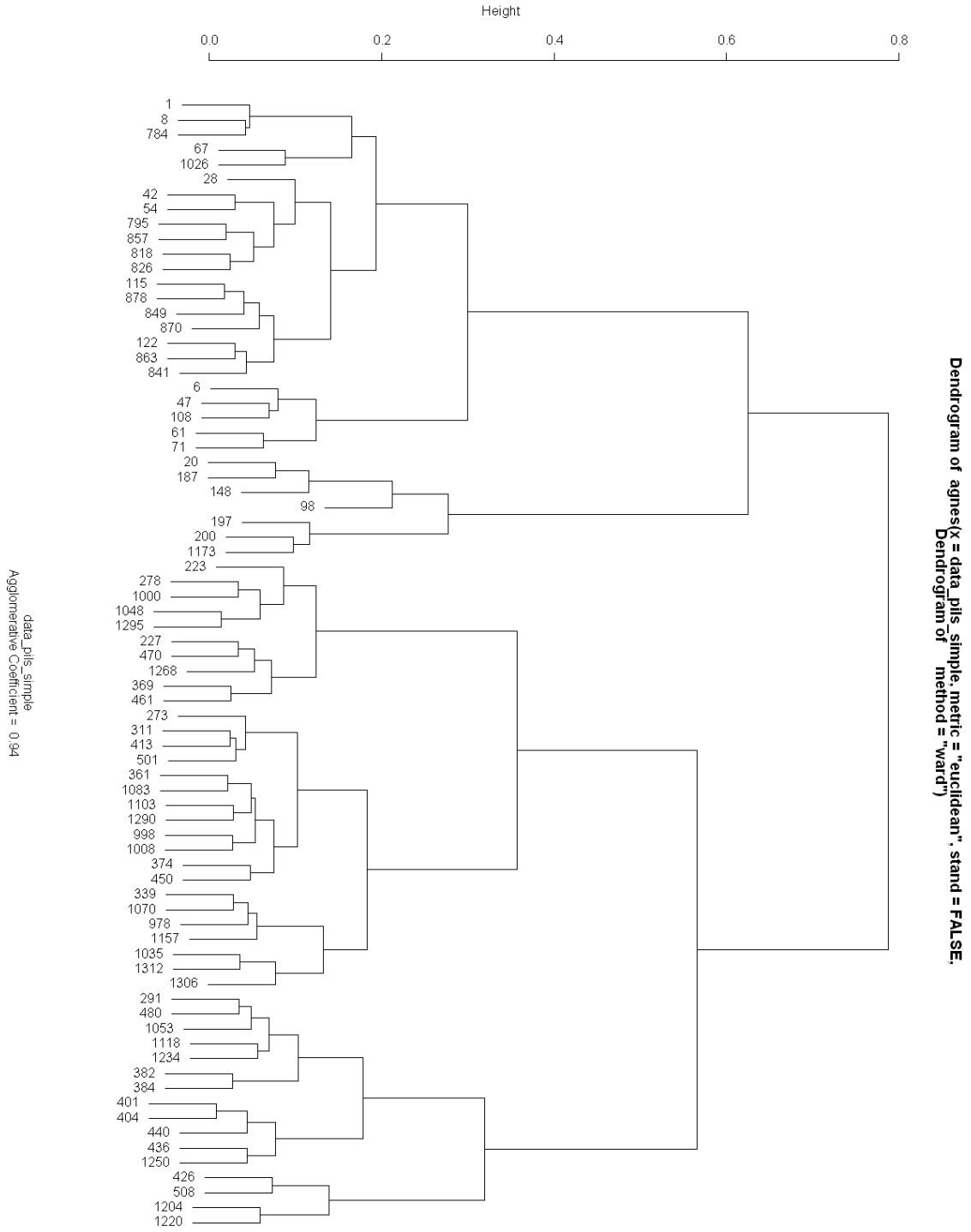


Figure B.17: Dendogram of Filtration Data from "Pilsener" using AGNES and Ward's Method

Table B.1: Evaluation of Clusters of *Lager B*

	binary label	**rating**	**describtion**
1. cluster	11100	3	poorer
2. cluster	10001	2	better
3. cluster	01101	3	poorer
4. cluster	01100	2	better
5. cluster	01101	3	poorer
6. cluster	00010	1	better
7. cluster	00010	1	better
8. cluster	11111	5	poorer
9. cluster	10110	3	poorer
10. cluster	10000	1	better

Table B.2: Evaluation of Clusters of *Nonalcoholic*

	binary label	**rating**	**describtion**
1. cluster	01011	3	poorer
2. cluster	00010	2	better
3. cluster	00001	1	better
4. cluster	11101	4	poorer
5. cluster	10000	1	better

B.2 Evaluation of Clusters

B.3 Interpretation of Clusters

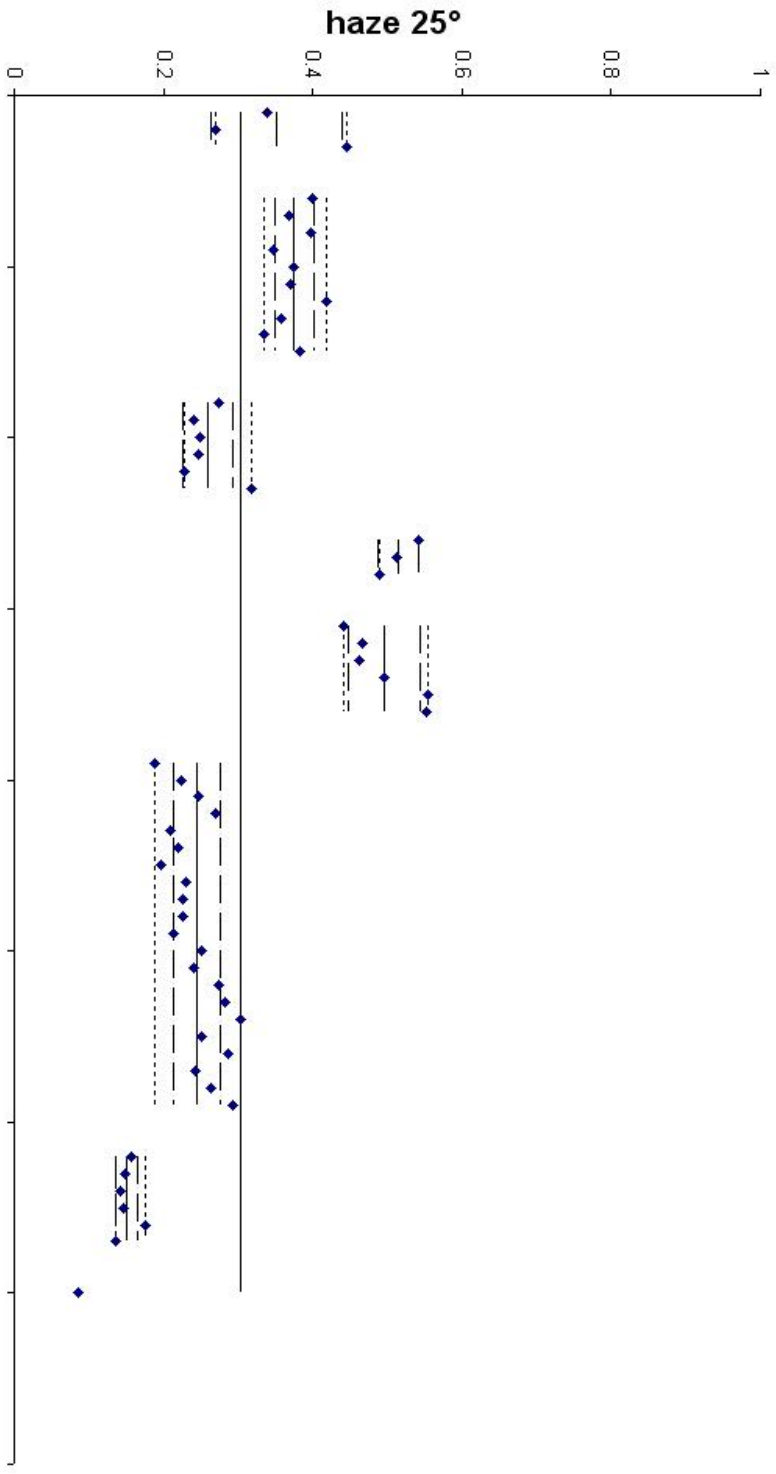


Figure B.18: Comparison of Means of $\bar{h}_{25}$

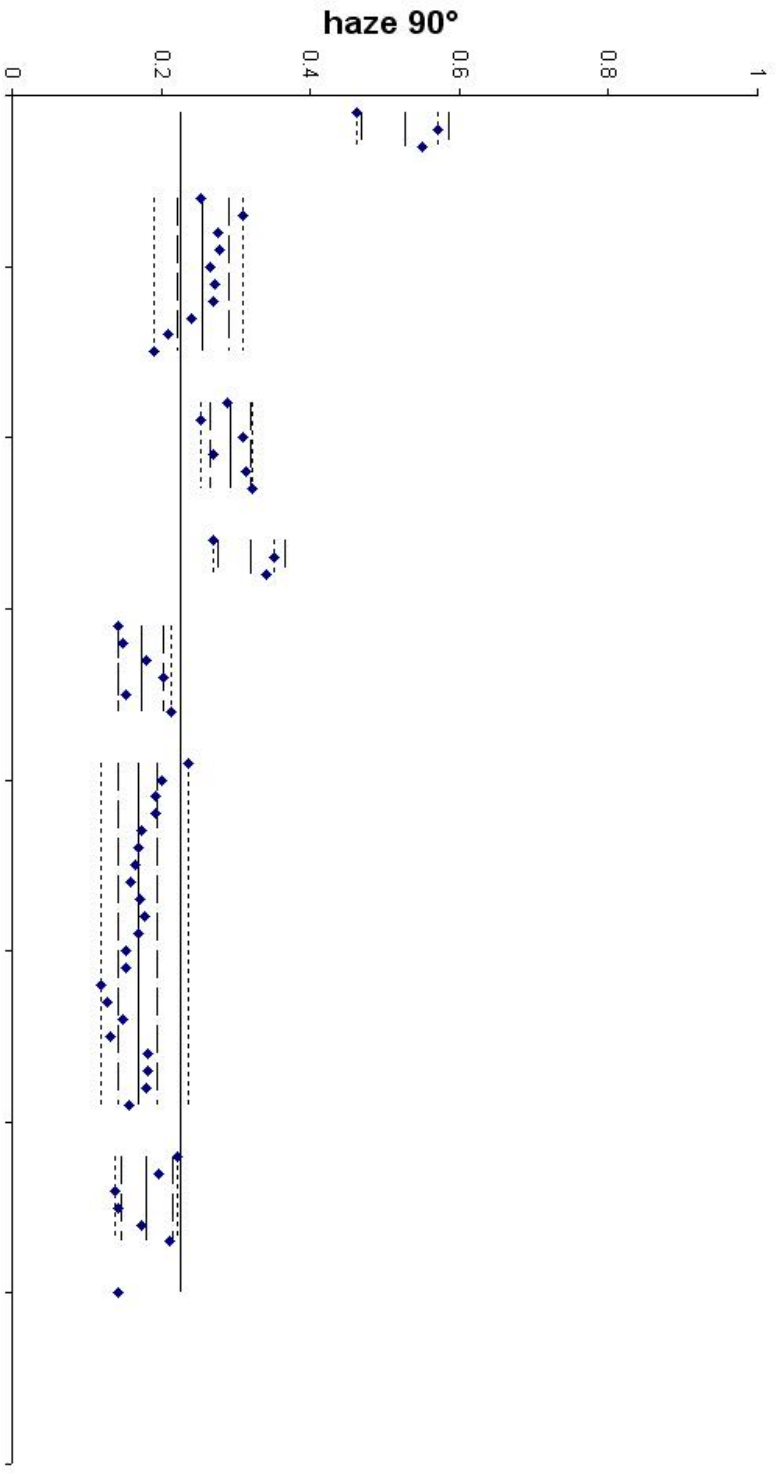


Figure B.19: Comparison of Means of $\bar{h}_{90}$

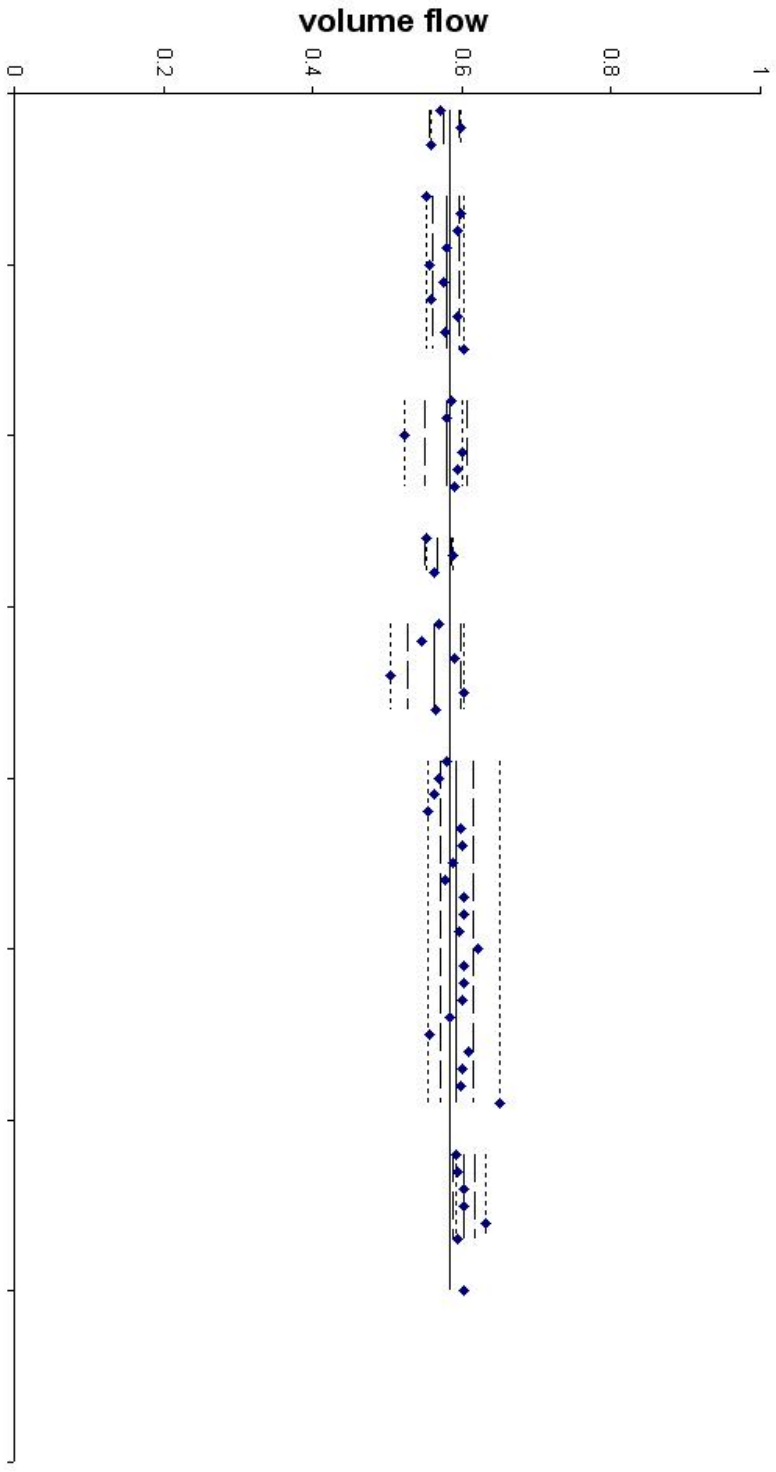


Figure B.20: Comparison of Means of $\bar{\dot{V}}$

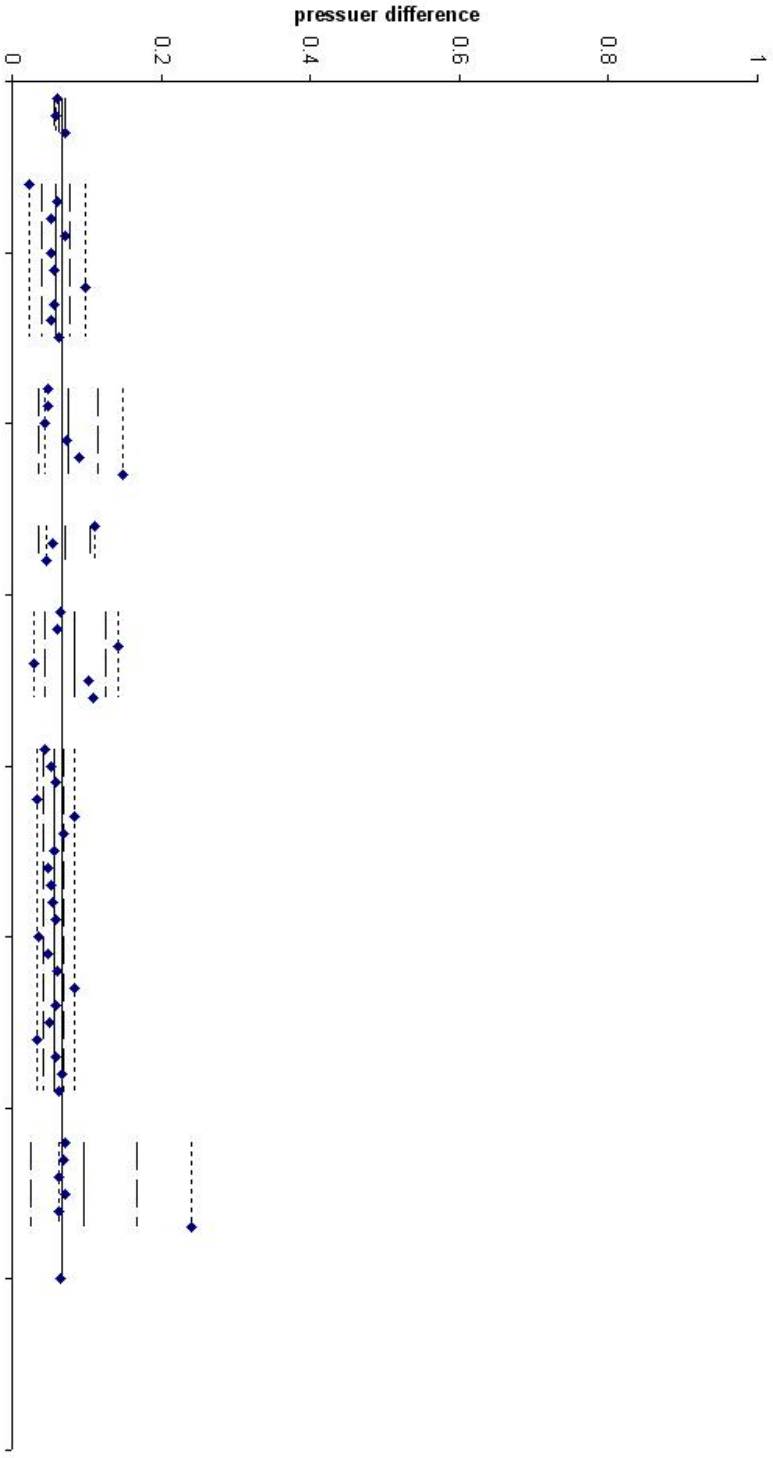


Figure B.21: Comparison of Means of Δp^*

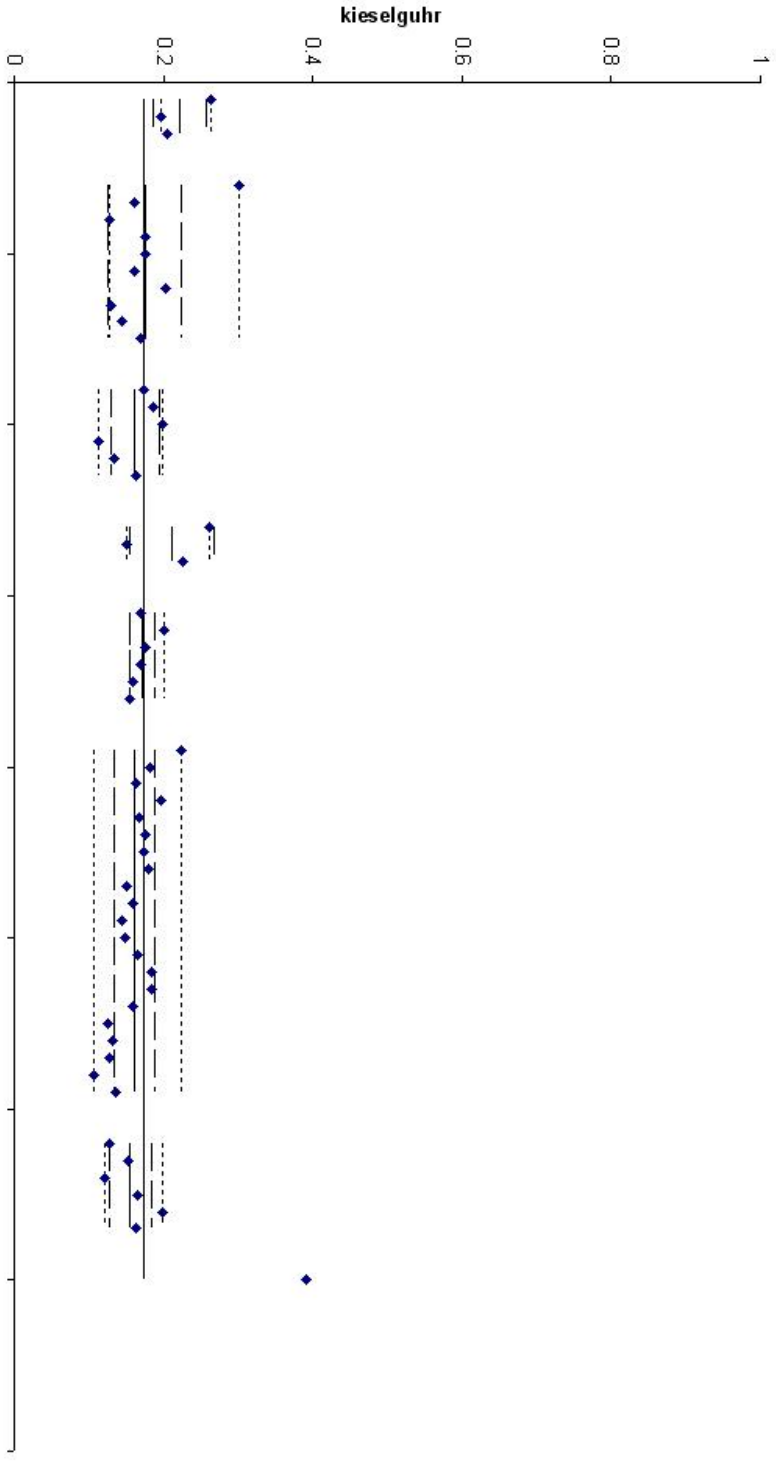


Figure B.22: Comparison of Means of $\sum KG$

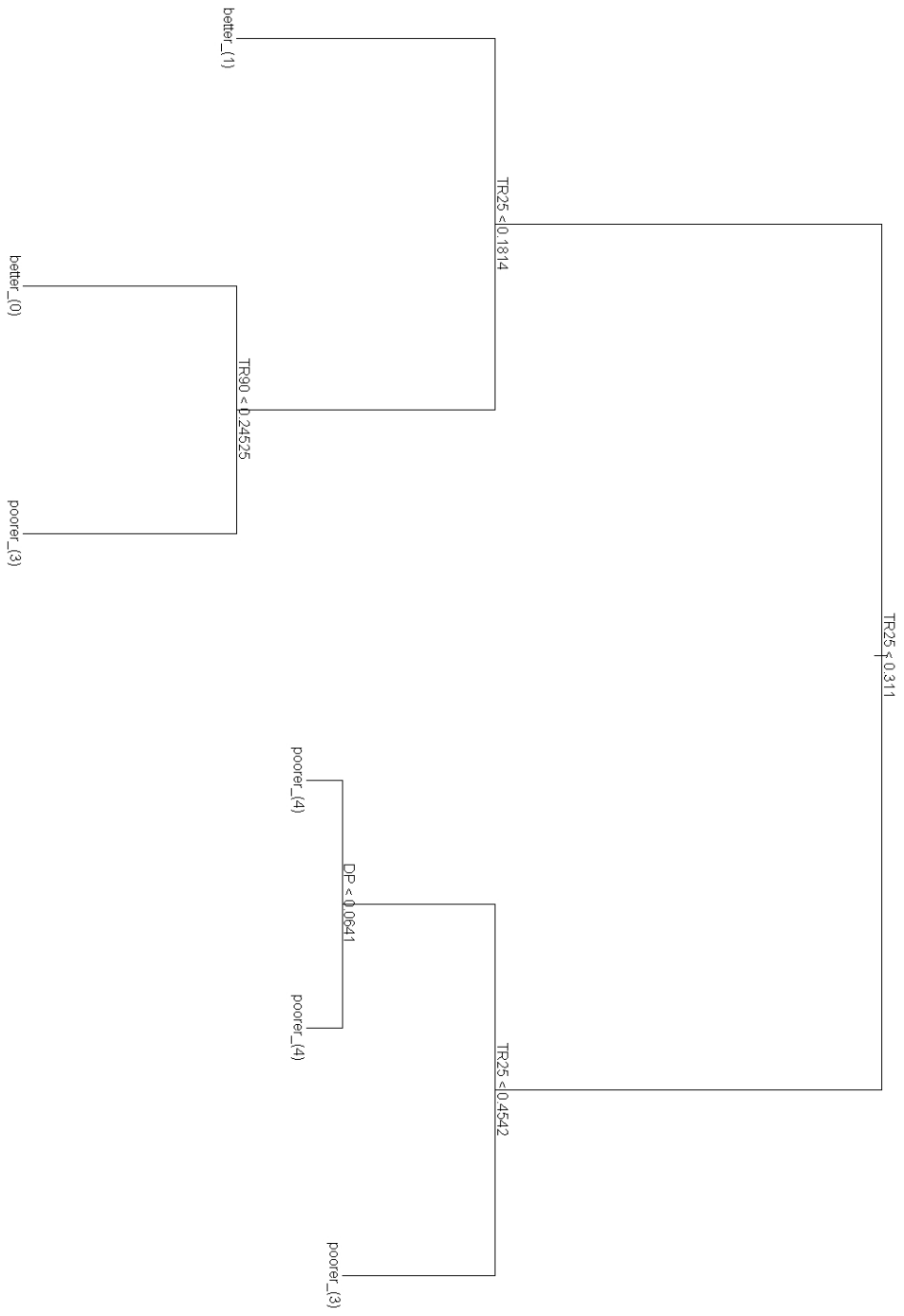


Figure B.23: Decision Tree of Filtration Data of *Lager A*

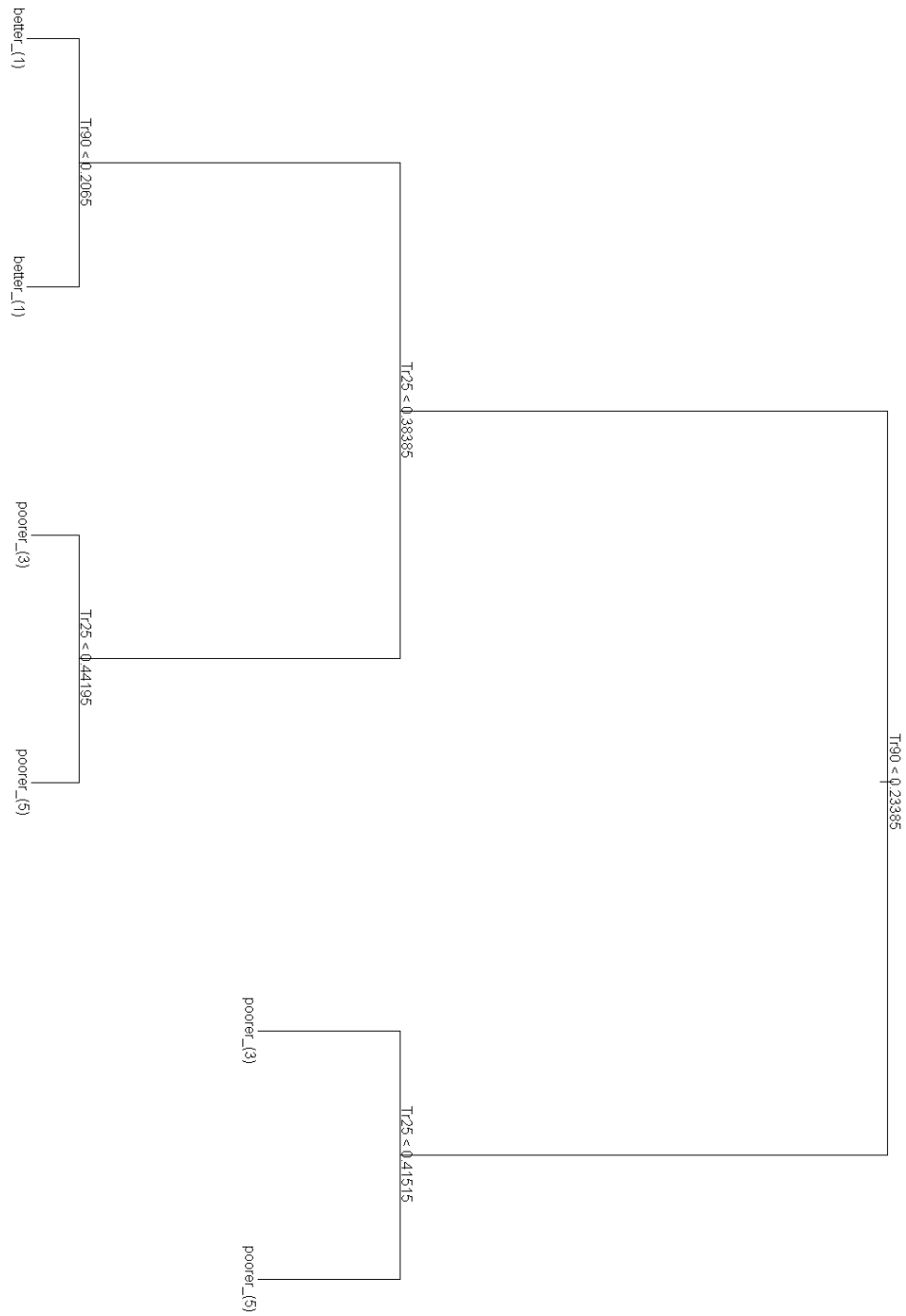


Figure B.24: Decision Tree of Filtration Data of *Lager B*

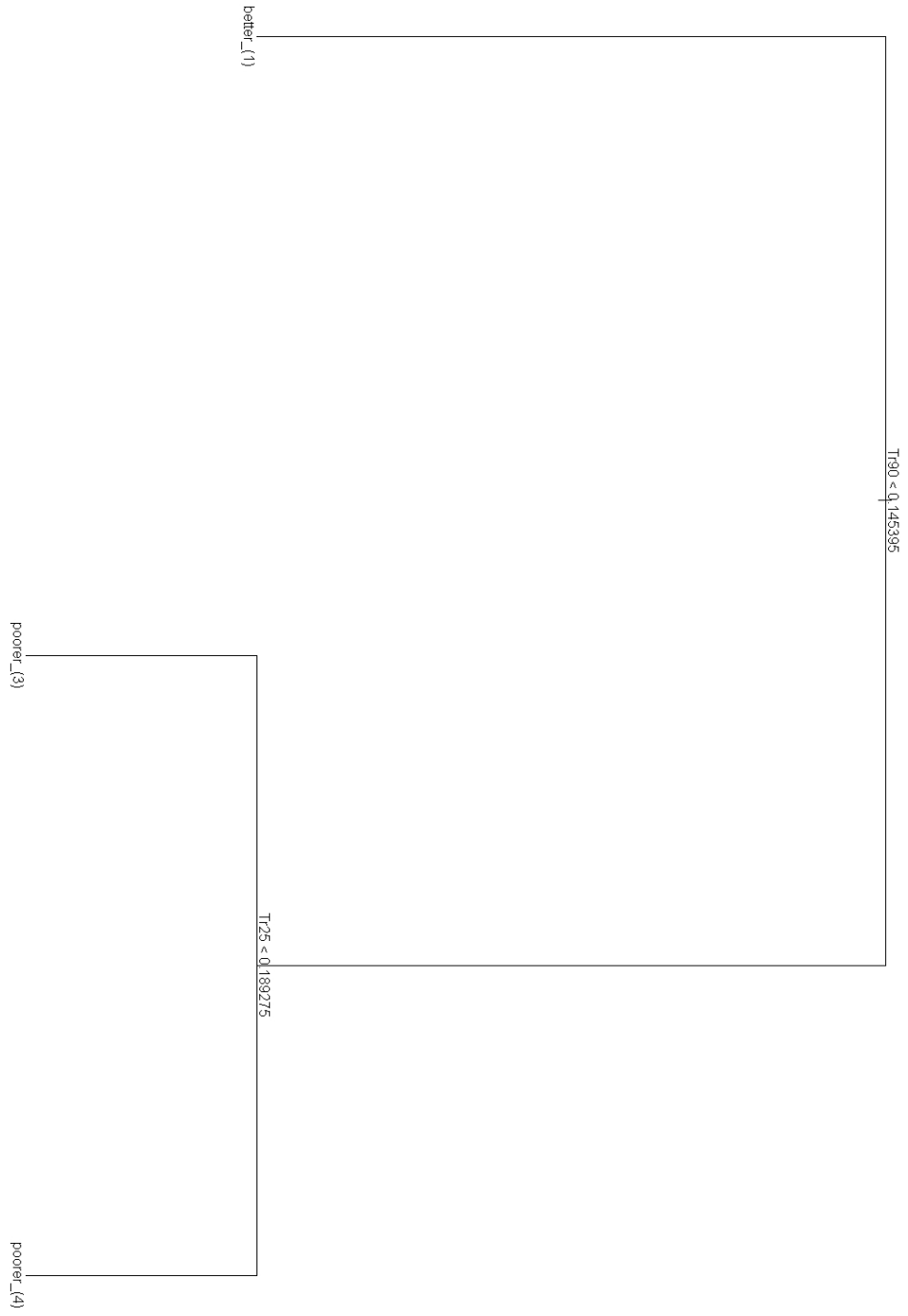


Figure B.25: Decision Tree of Filtration Data of *Nonalcoholic*

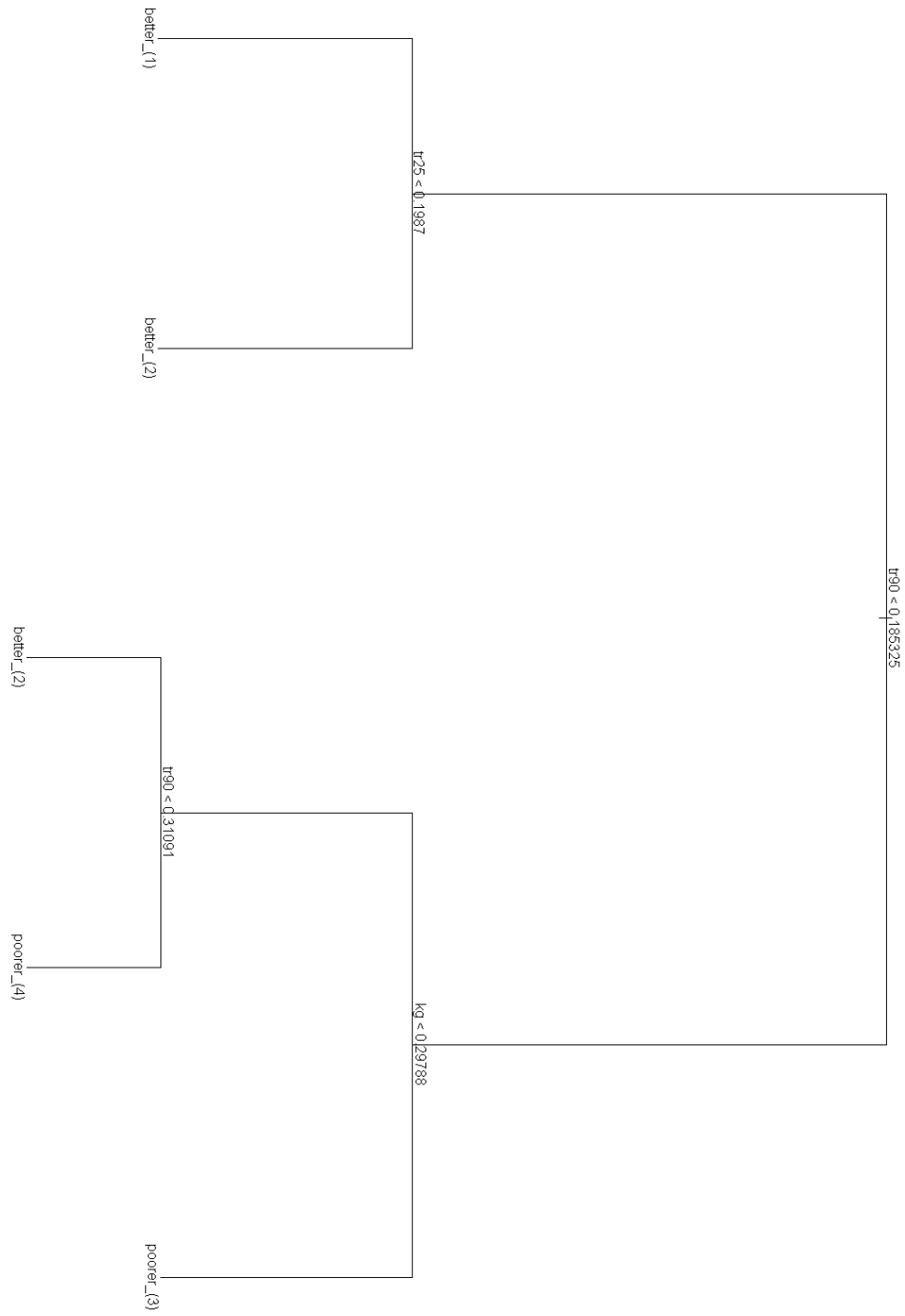


Figure B.26: Decision Tree of Filtration Data of *Pils*

Table B.3: Evaluation of Clusters of *Pils*

	binary label	**rating**	**describtion**
1. cluster	01100	2	better
2. cluster	01101	3	poorer
3. cluster	11111	5	poorer
4. cluster	11110	4	poorer
5. cluster	00001	1	better
6. cluster	00010	0	better
7. cluster	10001	2	better
8. cluster	10010	2	better

Appendix C

Classification of Filtration - Second Approach

C.1 Data Reduction

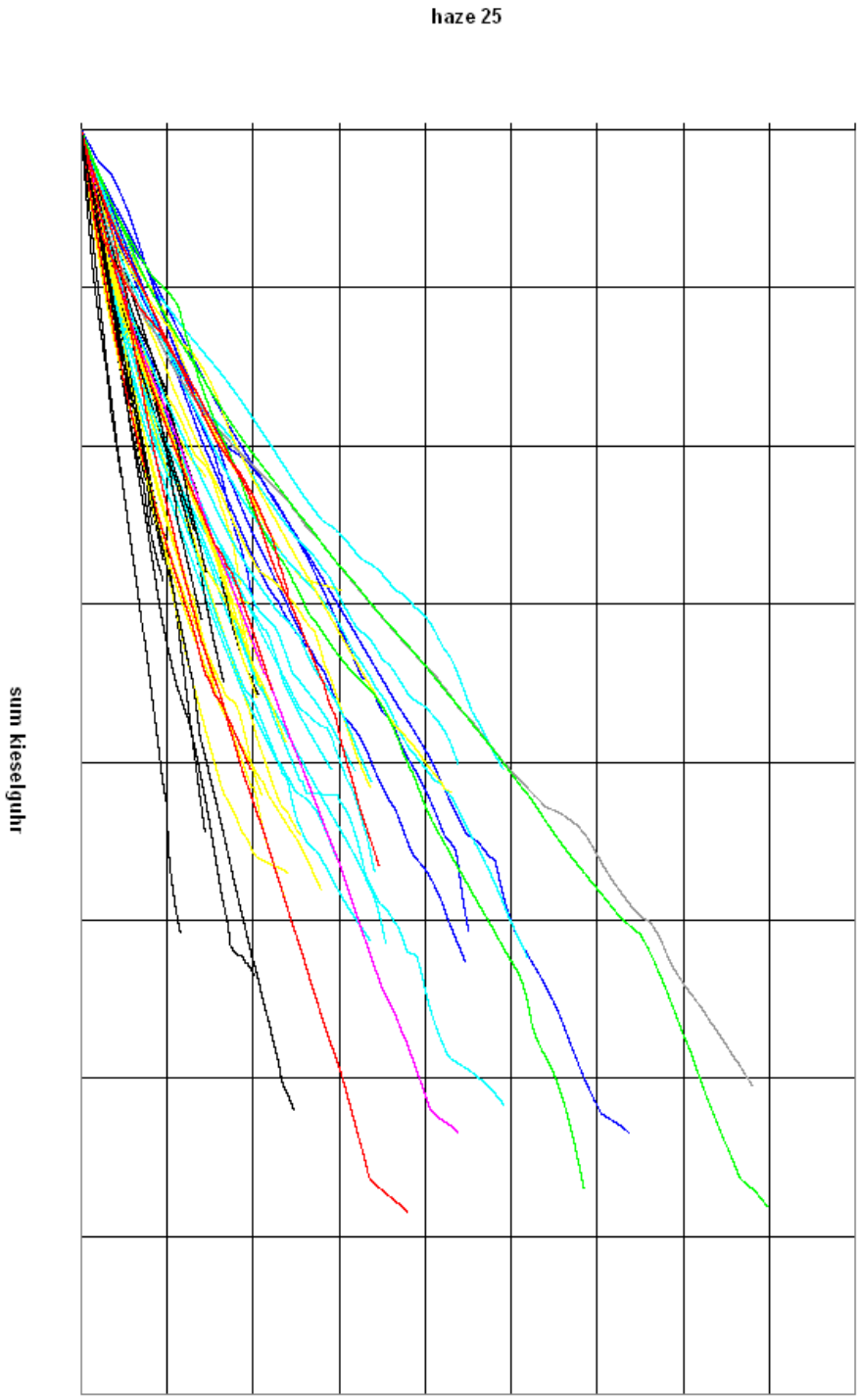


Figure C.1: Example Filtrations: Summation of h_{25} over $\sum KG$

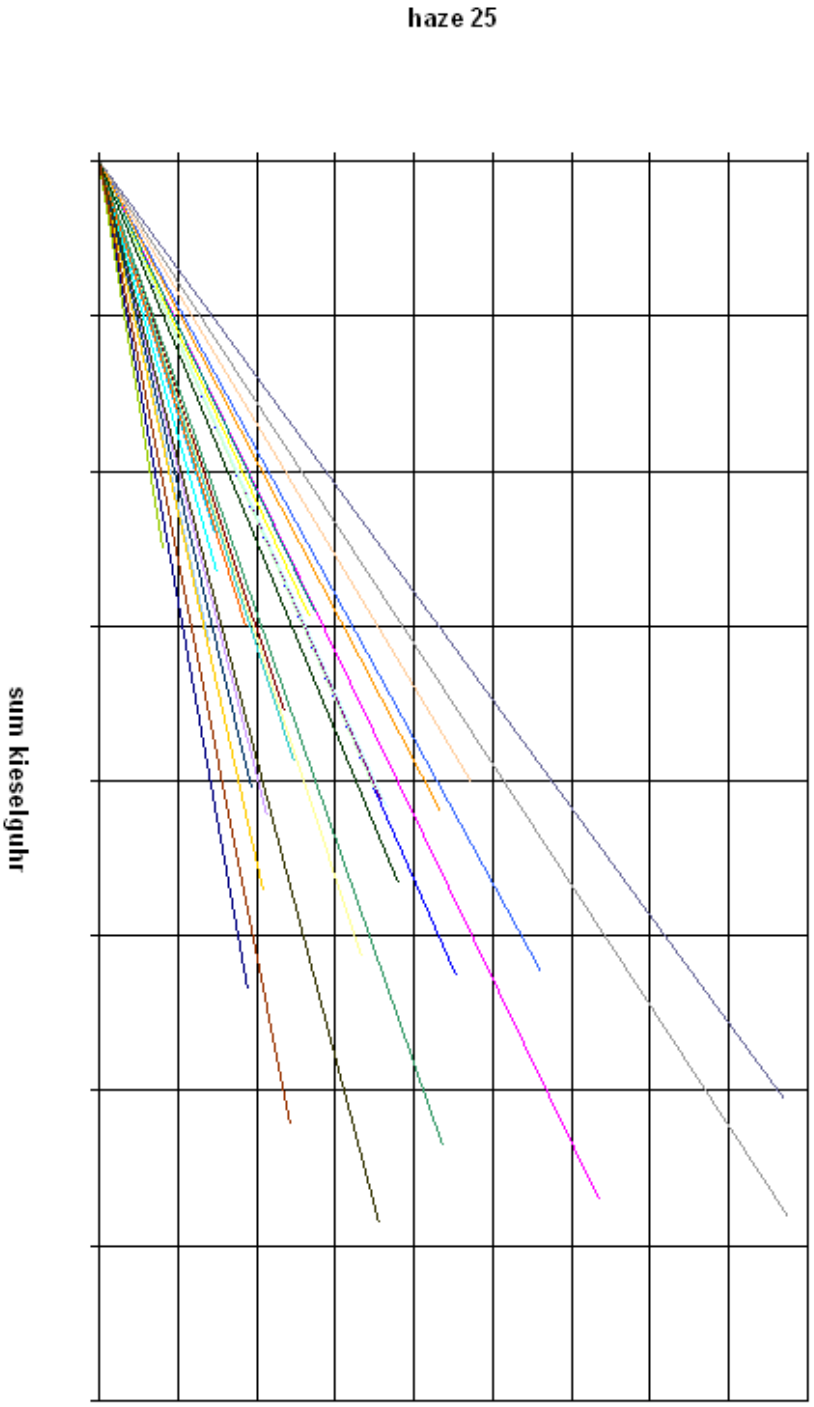


Figure C.2: Example Filtrations: Linear Regression of h_{25} over $\sum KG$

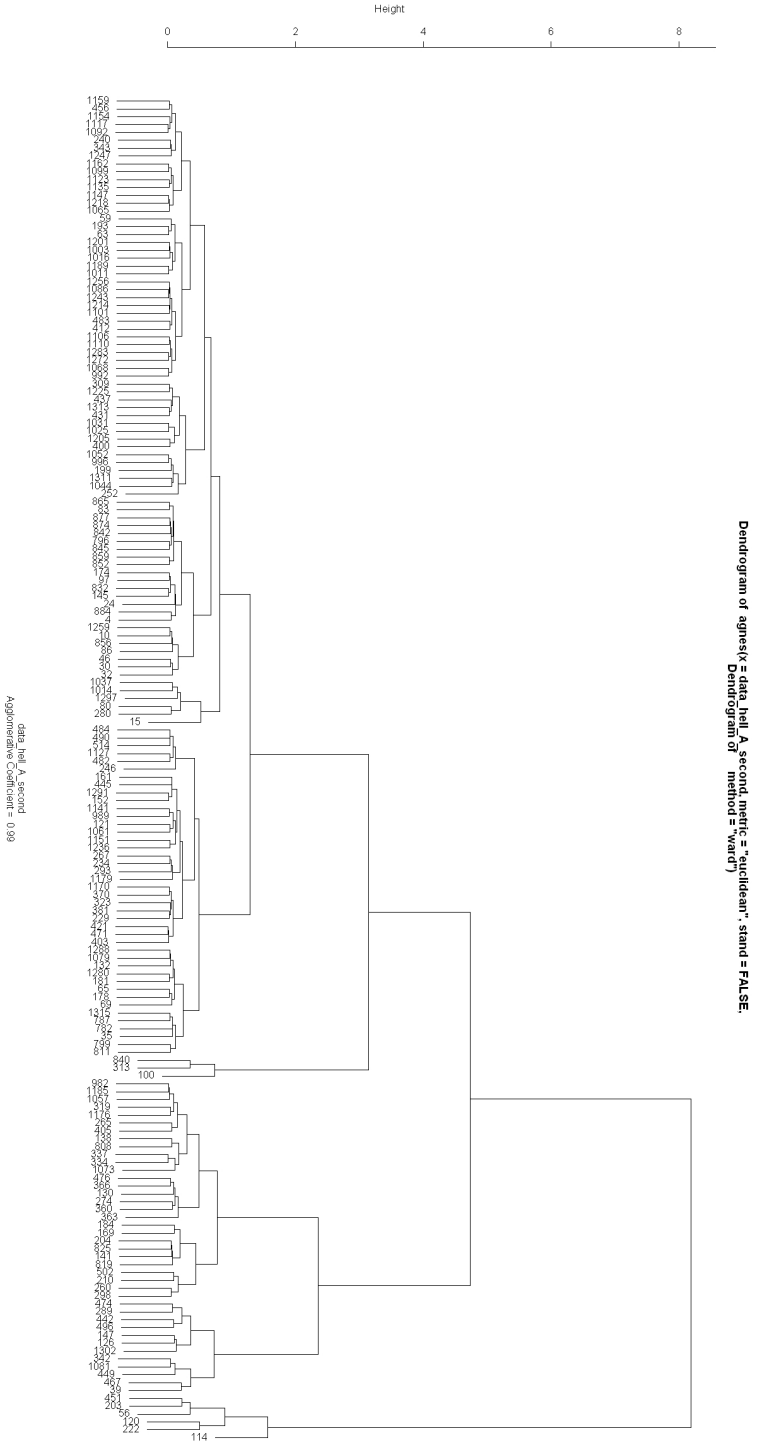


Figure C.3: Clustering of *hell A* filtration using the gradients of linear regression

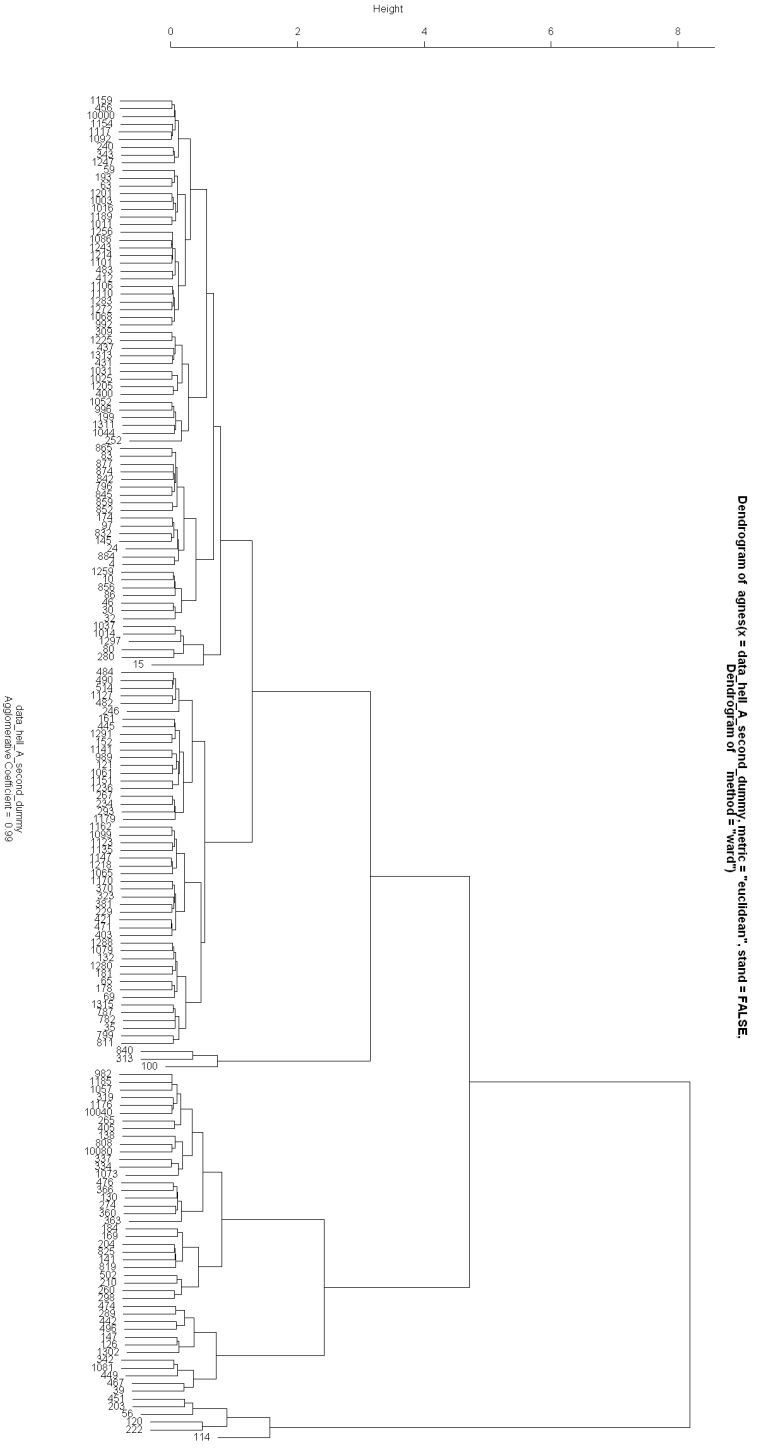


Figure C.4: Clustering of *hell A* filtration including dummies using the gradients of linear regression

```
K-means clustering with 3 clusters of sizes 34, 6, 134

Cluster means:
  tr25slope tr90slope dpvolslope
1 0.1291808 0.1766038  0.9145693
2 0.1556313 0.2296598  2.8016687
3 0.1413133 0.2048209  0.2643176

Clustering vector:
 1159   982   240    59  1201  1037   343    80  1256  1106
    3     3     3     3     3     3     3     3     3     3
  484   161  1283  1141   865   184  1162  1003   267  1214
    3     3     3     3     3     1     3     3     3     3
 1068   456    83  1259  1110   490   174  1288  1147   877
    3     3     3     3     3     3     3     3     3     3
  193  1170  1016   309    65  1225  1086   476    86  1272
    3     3     3     3     3     3     3     1     3     3
 1123   832   178  1154   884   204  1189  1031   323    69
    3     3     3     3     3     1     3     3     3     3
 1243  1101   483   121  1280  1135   859   181    15  1179
    3     3     3     3     3     3     3     3     3     3
 1011   474   265    46  1291  1052   796   342   132  1315
    3     1     3     3     3     3     3     1     3     3
 1079   840   405   147  1127   874   451   234    24  1205
    3     3     3     1     3     3     2     3     3     3
 1014   502   274    97  1302  1057   799   366   138  1081
    3     1     1     3     1     3     3     1     1     1
  842   421   169  1151   989   467   246    32  1218  1025
    3     3     1     3     3     1     3     3     3     3
  782   319   120  1311  1061   811   370   141  1099   845
    3     3     2     3     3     3     3     1     3     3
  437   203     4  1176   996   471   252    35  1247  1044
    3     2     3     3     3     3     3     3     3     3
  787   337   126  1313  1073   819   381   145  1117   852
    3     1     1     3     1     1     3     3     3     3
  449   222   442   210   514   289   992   334    56  1236
    1     2     1     1     3     1     3     1     2     3
  403   130   445   229   808   293    10  1065   360    63
    3     1     3     3     1     3     3     3     1     3
 1297   412   152   482   260   825   298    30  1092   363
    3     3     3     3     1     1     1     3     3     1
  100   431   199   496   280   856   313    39  1185   400
    3     3     3     1     3     3     3     1     3     3
  114 10000 10040 10080
    2     3     3     1
```

Figure C.5: *K-means* algorithm using random centroids

```
K-means clustering with 3 clusters of sizes 134, 34, 6

Cluster means:
  tr25slope tr90slope dpvolslope
1 0.1413133 0.2048209  0.2643176
2 0.1291808 0.1766038  0.9145693
3 0.1556313 0.2296598  2.8016687

Clustering vector:
 1159   982   240    59  1201  1037   343    80  1256  1106
    1     1     1     1     1     1     1     1     1     1
  484   161  1283  1141   865   184  1162  1003   267  1214
    1     1     1     1     1     2     1     1     1     1
 1068   456    83  1259  1110   490   174  1288  1147   877
    1     1     1     1     1     1     1     1     1     1
  193  1170  1016   309    65  1225  1086   476    86  1272
    1     1     1     1     1     1     1     2     1     1
 1123   832   178  1154   884   204  1189  1031   323    69
    1     1     1     1     1     2     1     1     1     1
 1243  1101   483   121  1280  1135   859   181    15  1179
    1     1     1     1     1     1     1     1     1     1
 1011   474   265    46  1291  1052   796   342   132  1315
    1     2     1     1     1     1     1     2     1     1
 1079   840   405   147  1127   874   451   234    24  1205
    1     1     1     2     1     1     3     1     1     1
 1014   502   274    97  1302  1057   799   366   138  1081
    1     2     2     1     2     1     1     2     2     2
  842   421   169  1151   989   467   246    32  1218  1025
    1     1     2     1     1     2     1     1     1     1
  782   319   120  1311  1061   811   370   141  1099   845
    1     1     3     1     1     1     1     2     1     1
  437   203     4  1176   996   471   252    35  1247  1044
    1     3     1     1     1     1     1     1     1     1
  787   337   126  1313  1073   819   381   145  1117   852
    1     2     2     1     2     2     1     1     1     1
  449   222   442   210   514   289   992   334    56  1236
    2     3     2     2     1     2     1     2     3     1
  403   130   445   229   808   293    10  1065   360    63
    1     2     1     1     2     1     1     1     2     1
 1297   412   152   482   260   825   298    30  1092   363
    1     1     1     1     2     2     2     1     1     2
  100   431   199   496   280   856   313    39  1185   400
    1     1     1     2     1     1     1     2     1     1
  114 10000 10040 10080
    3     1     1     2
```

Figure C.6: *K-means* algorithm using dummy filtration as centroids

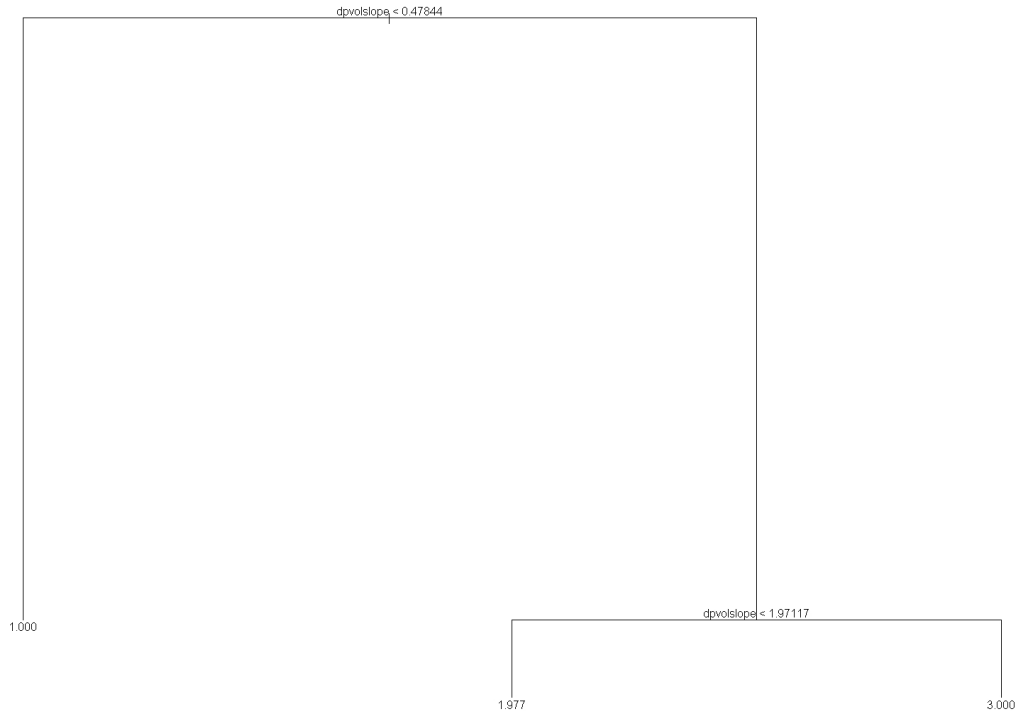


Figure C.7: Decision tree of *AGNES* cluster results

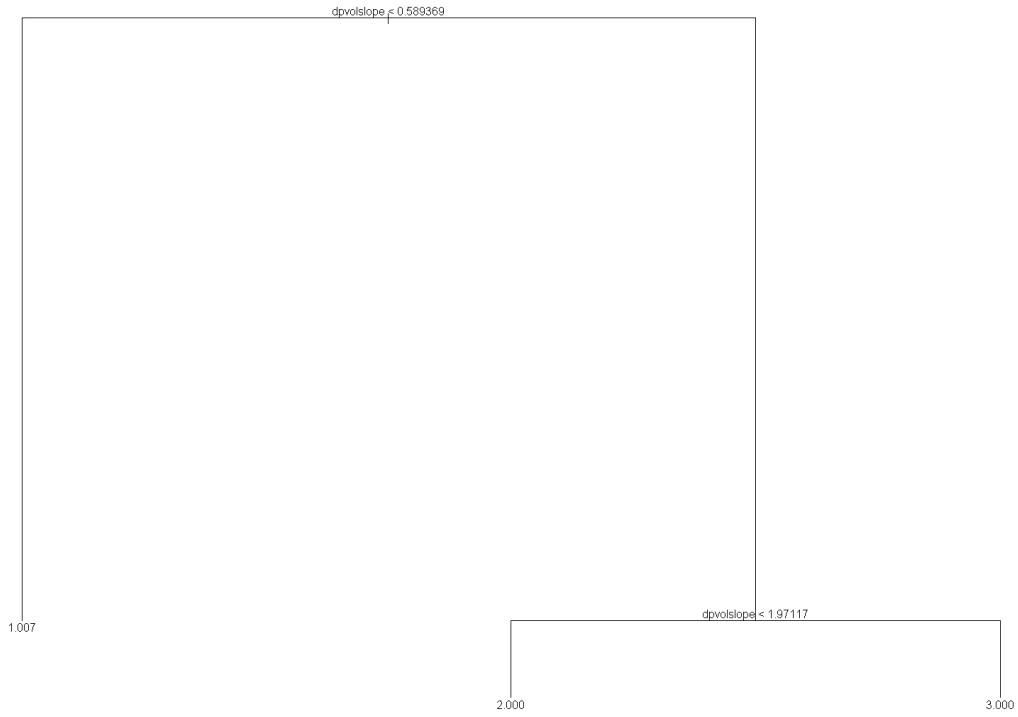


Figure C.8: Decision tree of *k-means* cluster results

Appendix D

Knowledge Discovery

D.1 Decision Tree Analysis

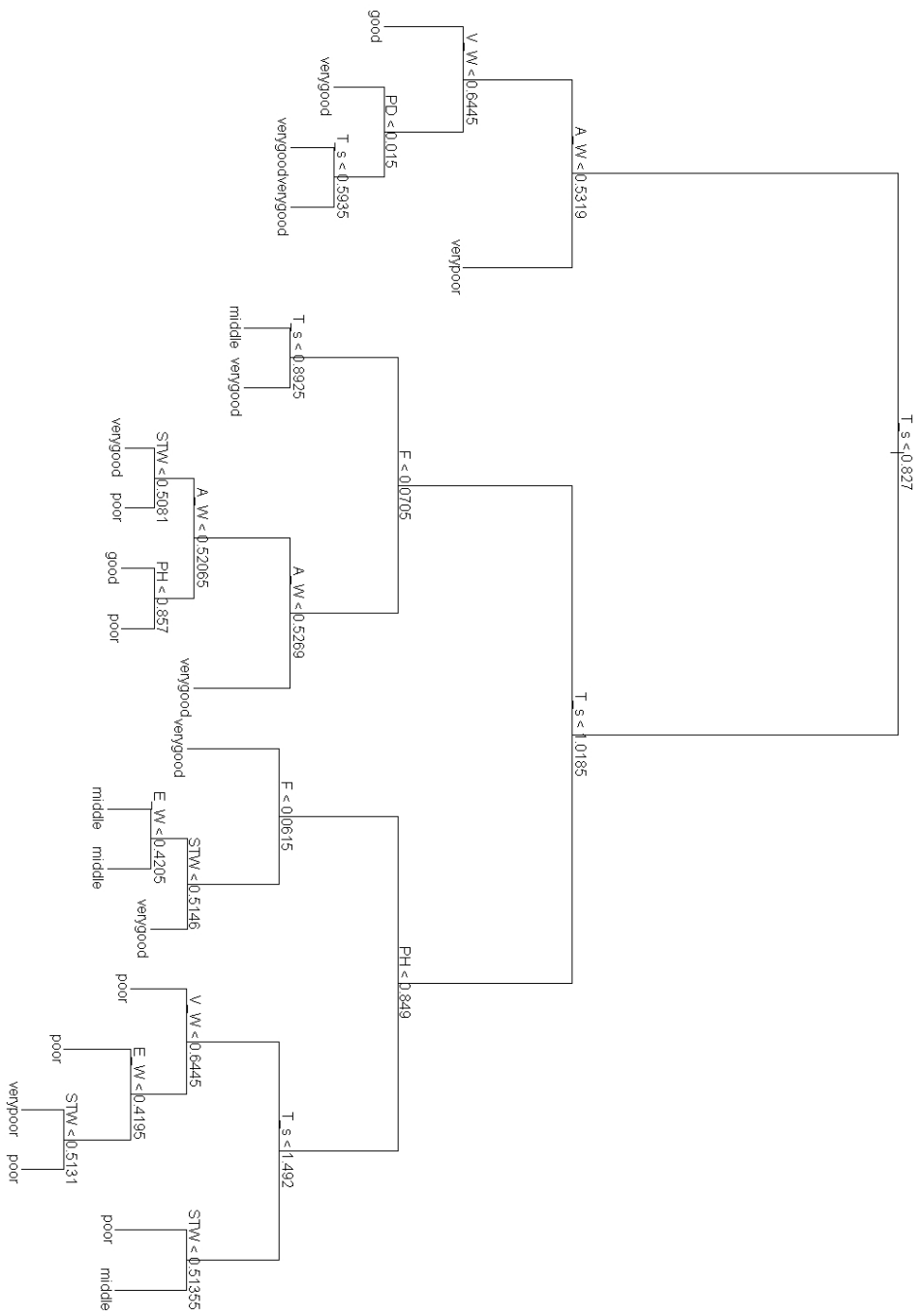


Figure D.1: Decision Tree of *lager A*

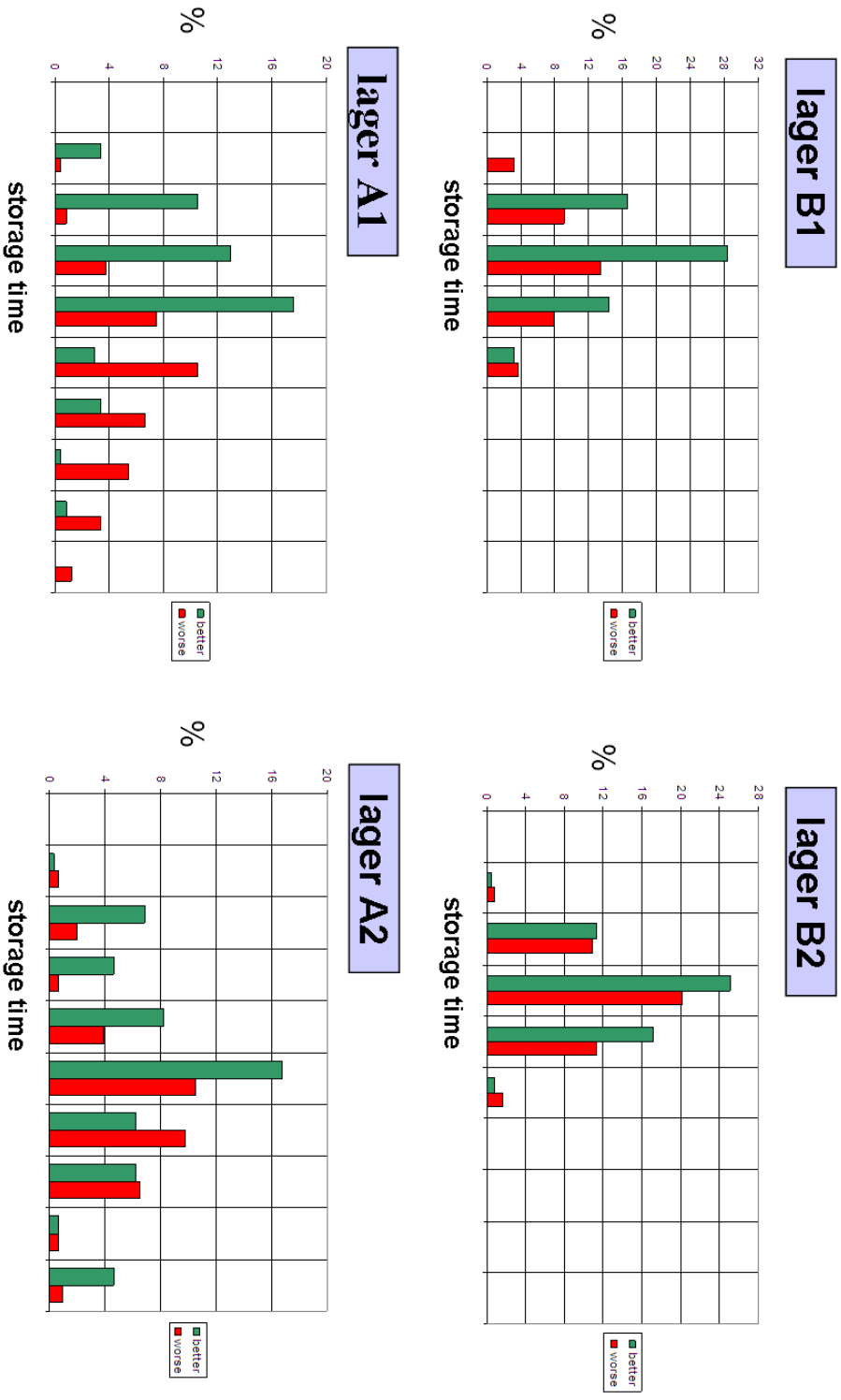


Figure D.2: Distribution of classes *better* and *worse* regarding storage time

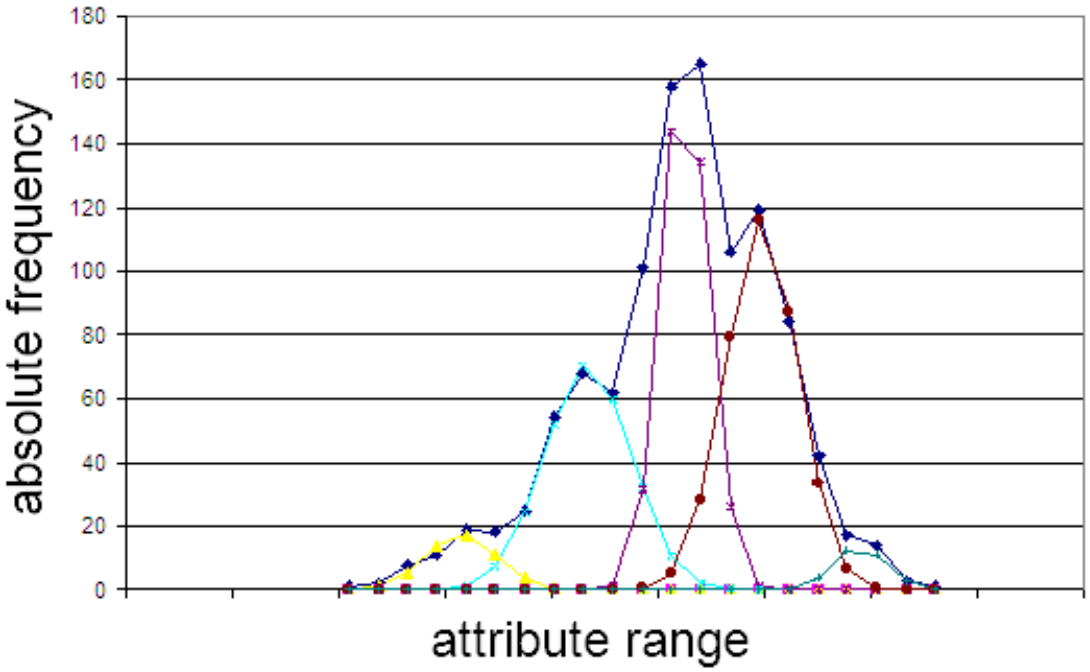


Figure D.3: Superposed Distributions of alcoholic content A_w of *lager A*

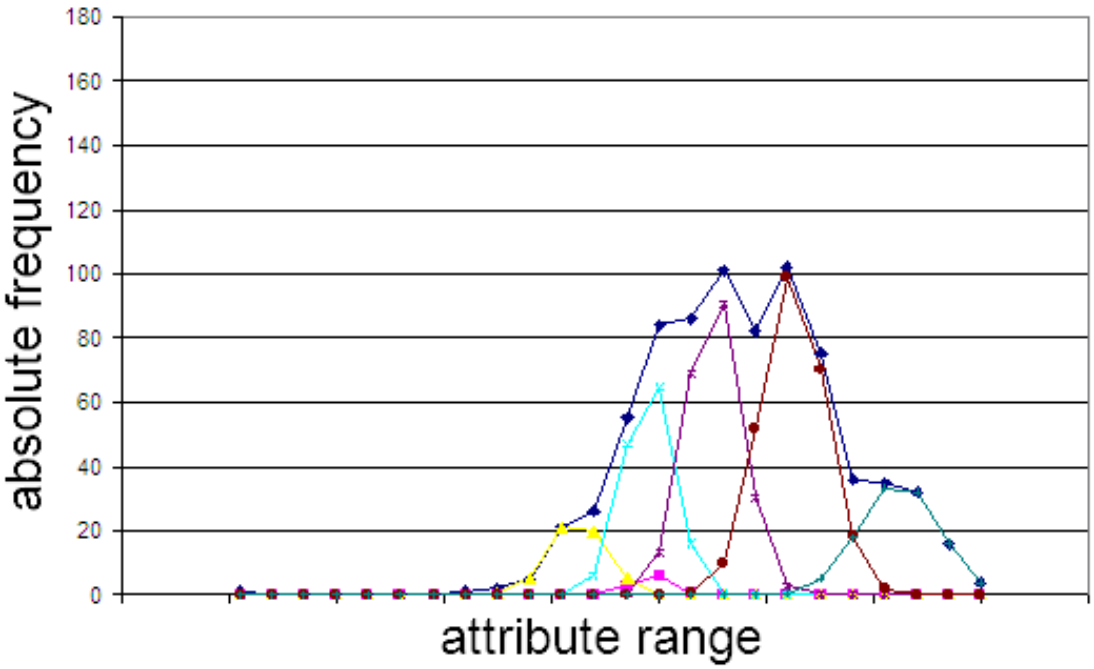


Figure D.4: Superposed Distributions of alcoholic content A_w of *lager B*

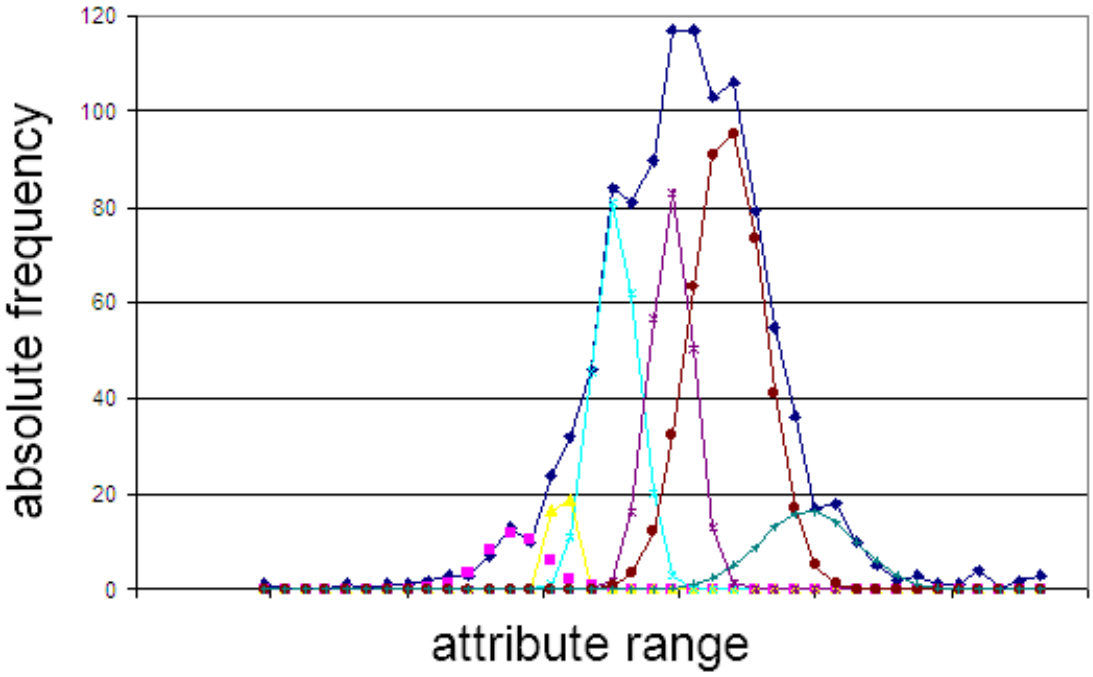


Figure D.5: Superposed Distributions of alcoholic content *STW* of *lager A*

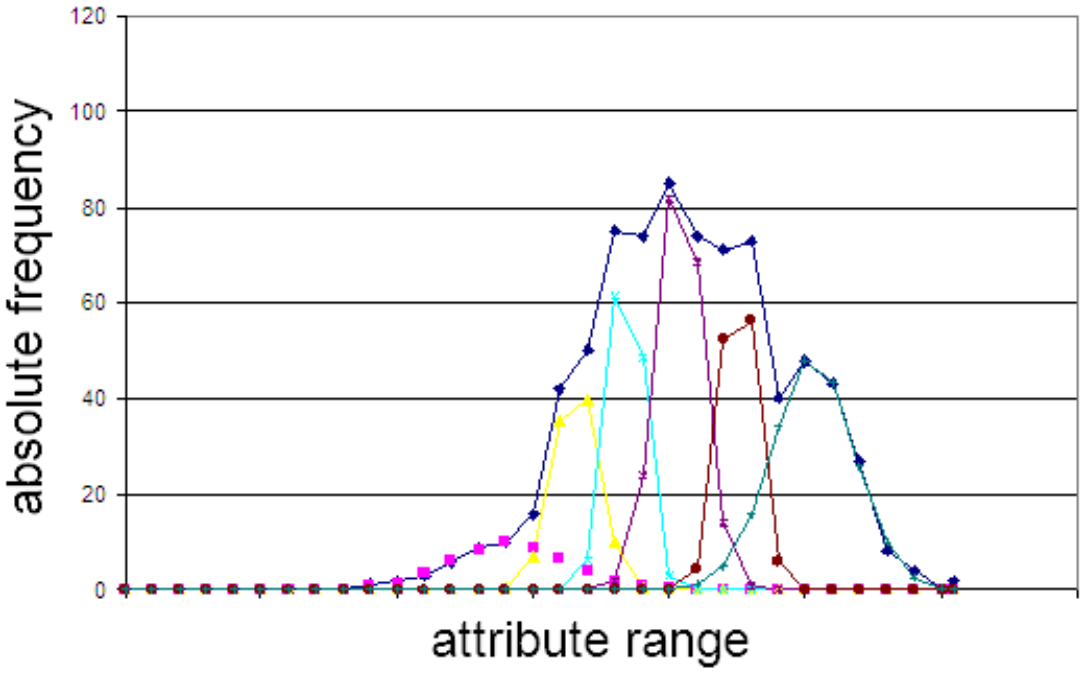


Figure D.6: Superposed Distributions of alcoholic content *STW* of *lager B*

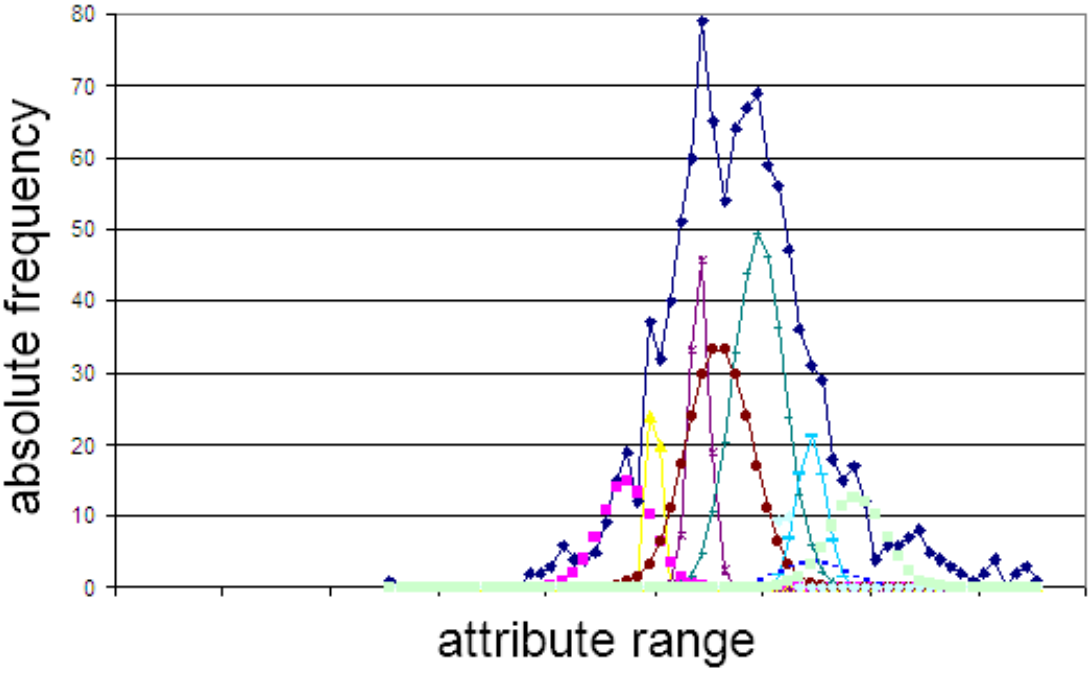


Figure D.7: Superposed Distributions of alcoholic content E_w of *lager A*

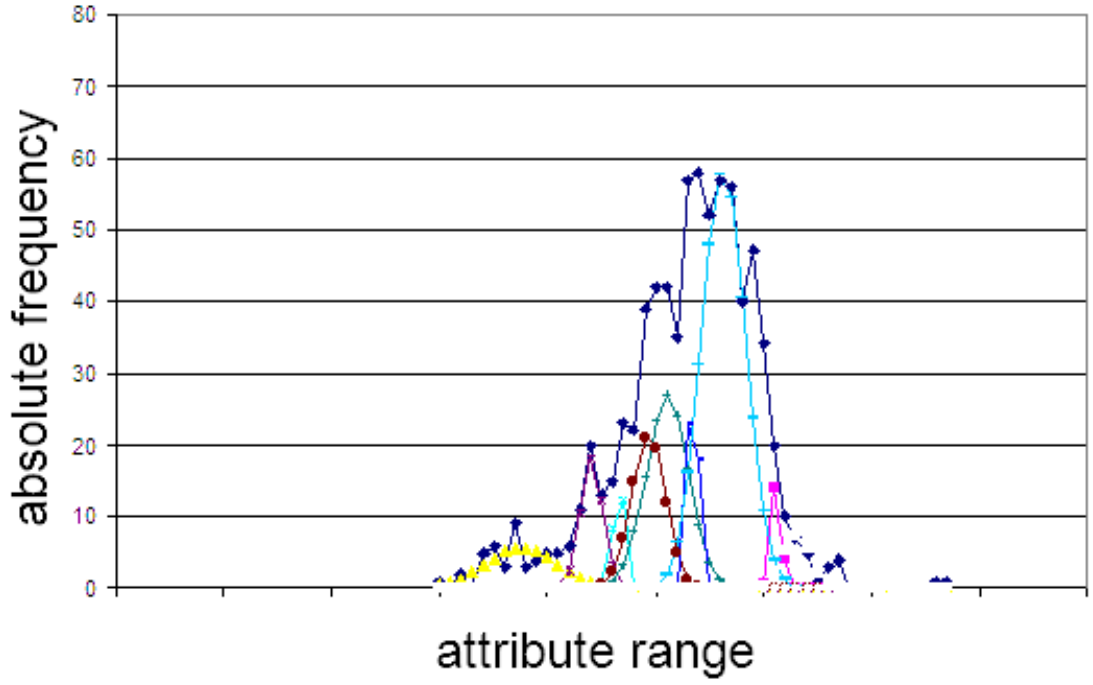


Figure D.8: Superposed Distributions of alcoholic content E_w of *lager B*

Michael Eberhard ♦ Haydstr. 17 ♦ 85354 Freising ♦ 08161 43444 ♦ eberhard.michael@krones.de

Lebenslauf

Persönliche Daten

Name:	Michael Eberhard
Geburtsdatum:	04.07.1976
Geburtsort:	Lage/Lippe
Staatsangehörigkeit:	deutsch
Familienstand:	ledig

Tätigkeiten

04/2006 -	Projekt-Ingenieur Buisness Intelligence Informationstechologie und Kommunikationstechnik (ITK) Krones AG / Neutraubling
02/2001 – 04/2006	Wissenschaftlicher Mitarbeiter / Stellvertretender Institutsleiter InformationsTechnologie Weihenstephan (ITW) des Wissenschaftszentrums Weihenstephan (WZW)

Ausbildung

11/1995 – 02/2001	Studium Brauwesen und Getränketechnologie (Diplom-Ingenieur), Technische Universität München / Weihenstephan Diplomarbeit: “Einfluß von Stabilisierungsmitteln auf die Filtrierbarkeit von Bier” Abschluß: Diplom-Ingenieur (Note: 2,1)
1986 – 1995	Gymnasium Lage Abschluß: Allgemeine Hochschulreife

Freising, 09.08.2006

Michael Eberhard