Rough Set Approach to Analyse Sicilian Milk Quality

Agata Matarazzo 1, Maria Teresa Clasadonte 1 and Nello Pappalardo 2

1 Department of Economics and Territory, University of Catania
   Corso Italia 55, 95129, Catania, Italy
   amatara@unict.it; clasadon@unict.it

2 Department of Agricultural Engineering, University of Catania
   Via Santa Sofia 100, 95123, Catania, Italy
   nello.pappalardo@unict.it

Abstract. This paper presents some important basic aspects of the classical Rough Sets approach applied to the values of chemical/microbiological parameters and information about the season and localization of the milk providers. By this approach, we have been able to study the “quality” of the milk in relation to several and inhomogeneous information, such as seasonal trends, place of origin, quantity of proteins, fat, somatic cells and total bacteria found in “high quality” milk samples, produced in a large area in southeast Sicily, during the years 2002-2004. This approach represents a mathematical tool quite different to classical statistics method able also to evaluate the importance of the considered variables. Besides, the results of this study are given in a clear and logical form (i.e. “if.. then...” decision rules) which highlight the relation between independent and dependent variables. Examples of such decision rules given by the most relevant condition attributes (independent variables), output of this approach, are represented in some tables, in order to show the most relevant features of the applied methodology. The data, collected every day and provided by a dairy industry, have been processed employing quantitative methodologies. The interesting results obtained with this new method encourage us to develop our research, studying other kinds of milk of different origin and quality. In: Sheibani K (ed). Proceedings of the 1st International Conference on Applied Operational Research – ICAOR (2008), pp 244–257. Lecture Notes in Management Science Vol. 1. ISSN 2008-0050.

Keywords: Rough sets, quality information, quantitative methods.

1 Introduction

The Italian dairy industry exceeds 18 billion euro per value of consumption, a turnover that offers a measurement of the importance of the driving sectors of our country’s agroindustry. In particular, in the last few years high quality milk has seen growth in sales of 6% and about 12% in terms of revenue, while its level of
penetration in Italian families has increased by 2.6% reaching 36 percentage points (Camasi B, 2004). Fresh pasteurized milk of high quality coming from secure and controlled farms, presents a habitual microbic endogenous count that is not very high; therefore, within 48 hours of the milking, it can be subject to a minimal thermic treatment, in this way conserving its vitaminic, proteinic and enzymatic properties almost unaltered, as is provided by law n. 169 del 3/5/1989 and by interdepartmental decree of 9/5/1991.

As is known, the qualitative characteristics of milk depend on multiple factors, some of which can be controlled at the origin by means of suitable prevention strategies (correct management of animal feeding, of their health conditions, of the udder’s hygiene, of the milking operations), while others can have an occasional nature and so be difficult to predict (U.C.P.L., 1996; Bonsembiante M., 2000). The directives issued by the European Union (Direttiva 92/46/CEE) on the full agreement to rigid hygienic-sanitary parameters set for the entire production chain show a top objective for safeguarding public health and guaranteeing the full functioning of the market to protect consumers’ economic interests. In particular, the D.M. n. 185 of 1991 (D.M. n. 185/91) and the D.M. n. 54/97 (D.M. n. 54/97) highlight how, in Italy, the distinction in qualitative classes of pasteurized milk is maintained in order to emphasize the objective of high quality in national productions (Desmasures N et al., 1997). The safeguarding of national production implies an indispensable improvement of hygienic-sanitary conditions and of the management of zootechnical companies, that finds against a cost of initial adjustment and maintenance for correct management, a slice of the market in growing expansion among those consumers who are more and more careful of nutritional food values and of its characteristics of freshness (Mazzaro A, 1999).

That being stated, in the current work, a study is being conducted intended to both individualize the seasonal and pedoclimatic influences on the product parameters that characterize the quality of the product, as well as to outline the qualitative profile of the companies producing raw milk of high quality, since their correct hygienic praxis surely depends on the total bacterial count (TBC), parameter taken into consideration by the Italian legislation for the distribution of qualitative classes of milk. In fact, even though the relationship of TBC with the microbiological quality of milk remains controversial (Park and Humphrey, 1986; Wilson et al., 1993; Zeng and Escobar, 1995), and TBC seems to always be influenced by various factors at the stage of lactation, parity, time of sampling (before, during or after milking) and stress (Haenlein, 2002; Sevi et al., 2004), and moreover the evaluation of factors affecting the total bacterial count in literature is rare (Zweifel et al., 2005). Comparable data analysing potential influence factors on the microbiological quality of raw small ruminant milk are rare in literature, and this study is the first dealing with this topic in Italy, with the advantage to analyse the milk quality taking into account also qualitative information such as the provider of the milk samples (the approach of the classical statistics used in the previous studies taking into account only quantitative data).

The study analyses data of chemical and microbiological parameters taken from samples of raw milk produced in Sicily that were daily conferred, in the period 2002-2004, to a dairy company in the province of Catania. The company, Zappala
s.r.l., that reaches the sixth national position in terms of brand in the mozzarella cheese area, provides, on a weekly basis, about 20000 points of sale distributed on the entire national territory, and utilizes for the production of “high quality milk” milk coming from farms distributed in the province of Ragusa constantly controlled by the local sanitary companies that verify the suitability for producing this food.

2 Data and methods

The classical Rough Set approach (Pawlak, 1991) is a mathematical tool used to handle vagueness and uncertainty inherent in making decisions. The concept of Rough Set theory is based on the assumption that every object of the universe (U) is associated with a certain amount of information (data, knowledge), expressed by means of some attributes (Q) used for object description. The most important problems that can be solved by the Rough Set theory are: finding a description of sets of objects in terms of attribute values, checking dependencies between attributes, reducing attributes and generating decision rules (Pawlak, 1997). Objects having the same description are indiscernible (similar) with respect to the available information. The indiscernibility relation thus generated constitutes a mathematical basis of the Rough Set theory; it induces a partition of the universe into blocks of indiscernible objects, called elementary sets, that can be used as parts to build knowledge about a real or abstract world. Any subset X of a universe may be expressed in terms of these elementary sets either precisely (as a union of elementary sets) or approximately only. In the latter case, the subset X may be characterized by two ordinary sets, called lower and upper approximations. The lower approximation of X is composed of all the elementary sets included in X (whose elements, therefore, certainly belong to X), while the upper approximation of X consists of all the elementary sets which have a non-empty intersection with X (whose elements, therefore, may belong to X).

Formally, let S be a decision table (used to represent knowledge), in which the attributes of the set Q are divided into condition attributes (set C≠Ø) and decision attributes (set D≠Ø), X a non-empty subset of U, P⊂Q a non-empty subset of attributes, the P-lower approximation and the P-upper approximation of X in S are defined, respectively, by (eqs. 1-2):

\[
P(X) = \{x \in U : I_p(x) \subseteq X\}
\]

\[
\bar{P}(X) = \bigcup_{x \in X} I_p(x)
\]

where \( I_p \) represents the indiscernibility relation on \( U \) with respect to a non-empty subset of attributes \( P \subseteq Q \) and \( I_p(x) = \{y \in U : y I_p x\} \) are the equivalence classes of \( x \in U \).
The difference between the upper and lower approximations constitutes the boundary region of the Rough Set, whose elements cannot be characterized with certainty as belonging or not to \( X \), using the available information. The information about objects from the boundary region is, therefore, inconsistent or ambiguous. For this reason, the number of objects of the boundary region may be used as a measure of vagueness of the information about \( X \). Moreover, the following ratio can be defined as the accuracy of the approximation of \( X \) (eq. 3), \( X \neq \emptyset \), by means of the attributes from \( P \):

\[
\alpha_P(X) = \frac{|P(X)|}{|X|}
\]  

(3)

The result is, obviously, \( 0 \leq \alpha_P(X) \leq 1 \); if \( \alpha_P(X)=1 \), \( X \) is an ordinary (exact) set with respect to \( P \); if \( \alpha_P(X)<1 \), \( X \) is a rough (vague) set with respect to \( P \).

It can also be defined a quality of the approximation of \( X \) by means of the attributes from \( P \) as (eq. 4):

\[
\gamma_P(X) = \frac{|P(X)|}{|X|}
\]  

(4)

The quality \( \gamma_P(X) \) represents the relative frequency of the objects correctly classified using the attributes from \( P \). Moreover, it results \( 0 \leq \gamma_P(X) \leq 1 \), \( \gamma_P(X)=0 \) iff \( \alpha_P(X)=0 \) and \( \gamma_P(X)=1 \) iff \( \alpha_P(X)=1 \).

The definition of approximations of a subset \( X \subseteq U \) can be extended to a classification, i.e. a partition \( Y = \{Y_1, \ldots, Y_n\} \) of \( U \). Subsets \( Y_i \), \( i=1, \ldots, n \), are disjunctive classes of \( Y \). By \( P \)-lower (\( P \)-upper) approximation of \( Y \) in \( S \) we mean sets \( P^Y = \{P_X \subseteq P_X \subseteq P_X \subseteq P_X \} \) and \( \overline{P}Y = \{\overline{P}_Y \subseteq \overline{P}_Y \subseteq \overline{P}_Y \subseteq \overline{P}_Y \} \), respectively. The coefficient of the eq. 5

\[
\gamma_P(Y) = \frac{\sum_{i=1}^{n} |PY_i|}{|U|}
\]

(5)

is called quality of the approximation of classification of \( Y \) by set of attributes \( P \), or in short, quality of classification. It expresses the ratio of all \( P \)-correctly classified objects to all objects in the system. Another important concept is that of
"superfluous" data in an information table. Superfluous data can be eliminated, in fact, without deteriorating the original information. Let $P \subseteq Q$ and $p \in P$. It is said that attribute $p$ is superfluous in $P$ if $l_{P} = l_{P \setminus \{p\}}$; otherwise, $p$ is indispensable in $P$. The set $P$ is independent if all its attributes are indispensable. The subset $P'$ of $P$ is a reduct of $P$ (denotation $Red(P)$) if $P'$ is independent and $l_{P} = l_{P'}$. A reduct of $P$ may also be defined with respect to an approximation of a partition $Y$ of $U$. It is then called $Y$-reduct of $P$ (denotation $Red_Y(P)$) and specifies a minimal subset $P'$ of $P$ which keeps the quality of classification unchanged, i.e. $\gamma_{P'}(Y) = \gamma_{P}(Y)$. In other words, the attributes that do not belong to $Y$-reduct of $P$ are superfluous with respect to the classification $Y$ of objects from $U$. The set containing all the indispensable attributes of $P$ is known as the $Y$-core (eq. 6). Formally:

$$Core_Y(P) = \bigcap Red_Y(P)$$ (6)

Since the $Y$-core, that could be equal to $O$, is the intersection of all the $Y$-reducts of $P$, it is included in every $Y$-reduct of $P$. It is the most important subset of attributes from $P$, because none of its elements can be removed without deteriorating the quality of classification.

Finally the Rough Set approach leads to the induction of a set of decision rules in the form of “if..., then...” statements, in order to represent the knowledge contained in the decision table. These rules represent the relations between a premise (if...), made of independent variables (condition attributes) and a consequence or decision (then...), expressed in terms of dependent variables (decision attributes). To select the most interesting rules the support and the confidence levels for each rule are evaluated. More precisely, the support of the rule represents the number of objects that match both the condition and the decision part of the rule. The confidence level expresses the ratio between the number of objects supporting the rule and the number of all objects matching the condition part of the rule.

The Rough Set approach based on the indiscernibility relation was applied to analyse and compare the different milk quality by means of several descriptors (delivery date of the milk to the farm, supplier which produces the raw material, and quantity of proteins, fat, somatic cells and bacteria present in the milk samples). The database regarding this information was provided by Mr Alfio Zappala’s Quality Control Laboratory of his dairy farm (Zafferana Etnea – Eastern Sicily). The information of the initial database was subdivided in two databases containing respectively the data of the analyses corresponding to “high quality” or “normal” milk (the analysed milk samples were provided for the period 2002 - 2004). Accounting of a high amount of data originating from the milk of extremely diversified farms cannot avoid taking into account the high variability existing as several factors (for instance race, type of stalling, feeding, etc.) on which hypotheses can be made and that will be the subject of further research.

The raw data of each considered variable, as required by the Rough Set method based on the indiscernibility relation, were discretised in 4 subsets. More precisely, the values sets of the variables (fat, proteins, somatic cells and total bacteria content) have been properly discretised in order to build “robust” classes of data also with respect to the tails of their distribution and obtain classes of equal
probability of occurrence (Barbagallo et al., 2006; Clasadonte et al., 2003, 2004). The results of the discretization are reported in Tables 1-2.

**Table 1.** Discretization of the attributes (milk samples of normal quality)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Class no.1</th>
<th>Class no.2</th>
<th>Class no.3</th>
<th>Class no.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery date (season)</td>
<td>Winter</td>
<td>Spring</td>
<td>Summer</td>
<td>Autumn</td>
</tr>
<tr>
<td>Provider</td>
<td>Each number corresponds to a firm which produces the raw material</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fat content (%)</td>
<td>[2.7-3.486]</td>
<td>[3.486-3.654]</td>
<td>[3.654-3.822]</td>
<td>[3.822-4.7]</td>
</tr>
<tr>
<td>Protein content (%)</td>
<td>[2.8-3.23]</td>
<td>[3.23-3.349]</td>
<td>[3.349-3.469]</td>
<td>[3.469-3.9]</td>
</tr>
<tr>
<td>Somatic cells content</td>
<td>[60-347.833]</td>
<td>[347.833-587.05]</td>
<td>[587.65-826.267]</td>
<td>(826.267-2.820]</td>
</tr>
<tr>
<td>(cells/mL*10^4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** Discretization of the attributes (milk samples of high quality)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Class no.1</th>
<th>Class no.2</th>
<th>Class no.3</th>
<th>Class no.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery date (season)</td>
<td>Winter</td>
<td>Spring</td>
<td>Summer</td>
<td>Autumn</td>
</tr>
<tr>
<td>Provider</td>
<td>Each number corresponds to a firm which produces the raw material</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fat content (%)</td>
<td>[2.67-3.48]</td>
<td>[3.48-3.60]</td>
<td>[3.60-3.72]</td>
<td>[3.72-4.05]</td>
</tr>
<tr>
<td>Protein content (%)</td>
<td>[3.08-3.26]</td>
<td>[3.26-3.328]</td>
<td>[3.328-3.395]</td>
<td>[3.395-3.67]</td>
</tr>
<tr>
<td>Somatic cells content</td>
<td>[56-214.578]</td>
<td>[214.578-324.498]</td>
<td>[324.498-434.418]</td>
<td>(434.418-1.041]</td>
</tr>
<tr>
<td>(cells/mL*10^4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total bacteria content (TBC)</td>
<td>[7-20.42]</td>
<td>(20.42-30.828]</td>
<td>(30.828-41.233]</td>
<td>(41.233-95]</td>
</tr>
</tbody>
</table>

Since some of these variables have also been considered in turn decision attributes, in these cases, in order to obtain very general rules and improve their quality and confidence level, the domain of these attributes was discretised in two intervals (classes). This kind of discretization is shown in Tables 3-4.

**Table 3.** Discretization of the decision attributes (milk samples of normal quality)

<table>
<thead>
<tr>
<th>Decision attribute</th>
<th>Decision class no.1</th>
<th>Decision class no.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat content (%)</td>
<td>[2.70-3.65]</td>
<td>[3.65-4.70]</td>
</tr>
<tr>
<td>Protein content (%)</td>
<td>[2.80-3.35]</td>
<td>[3.35-3.90]</td>
</tr>
<tr>
<td>Somatic cells content</td>
<td>[60.00-587.05]</td>
<td>[587.05-2.820]</td>
</tr>
<tr>
<td>(cells/mL*10^4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total bacteria content (TBC)</td>
<td>[1.00-116.21]</td>
<td>[116.21-2.500]</td>
</tr>
</tbody>
</table>

**Table 4.** Discretization of the decision attributes (milk samples of high quality)

<table>
<thead>
<tr>
<th>Decision attribute</th>
<th>Decision class no.1</th>
<th>Decision class no.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat content (%)</td>
<td>[2.67-3.60]</td>
<td>[3.60-4.05]</td>
</tr>
<tr>
<td>Protein content (%)</td>
<td>[3.08-3.33]</td>
<td>[3.33-3.67]</td>
</tr>
<tr>
<td>Somatic cells content</td>
<td>[56.00-324.50]</td>
<td>[324.50-1.041]</td>
</tr>
<tr>
<td>(cells/mL*10^4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total bacteria content (TBC)</td>
<td>[7.00-30.83]</td>
<td>[30.83-95.00]</td>
</tr>
</tbody>
</table>
3 Results and discussion

The software Rose (Predki et al., 1998, 1999) developed by IDSS (University of Poznan, Poland, 2000) was applied to eight decision tables (four for each kind of milk samples), each time using the fat, proteins, somatic cells and TBC as decision attributes (keeping the above mentioned information as condition attributes). The results of Rose application in terms of quality of classification, reducts and core are presented in Tables 5-6.

Table 5. Outputs of Rose application - analysis of the normal quality milk

<table>
<thead>
<tr>
<th>Decision attribute</th>
<th>Quality of Classification</th>
<th>Reducts</th>
<th>Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBC</td>
<td>0.86</td>
<td>Season, provider, somatic cells, protein</td>
<td>Season, provider, somatic cells, protein</td>
</tr>
<tr>
<td>Somatic cells content</td>
<td>0.82</td>
<td>Season, provider, fat, protein, TBC</td>
<td>Season, provider, fat, protein, TBC</td>
</tr>
<tr>
<td>Fat content</td>
<td>0.80</td>
<td>Season, provider, protein, TBC</td>
<td>Season, provider, protein</td>
</tr>
<tr>
<td>Protein content</td>
<td>0.85</td>
<td>Season, provider, fat, somatic cells, TBC</td>
<td>Season, provider, fat, somatic cells, TBC</td>
</tr>
</tbody>
</table>

Table 6. Outputs of Rose application - analysis of the high quality milk

<table>
<thead>
<tr>
<th>Decision attribute</th>
<th>Quality of Classification</th>
<th>Reducts</th>
<th>Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBC</td>
<td>0.82</td>
<td>Coinciding with all the present condition attributes</td>
<td>Identical to the reduct</td>
</tr>
<tr>
<td>Somatic cells content</td>
<td>0.80</td>
<td>Coinciding with all the present condition attributes</td>
<td>Identical to the reduct</td>
</tr>
<tr>
<td>Fat content</td>
<td>0.72</td>
<td>Coinciding with all the present condition attributes</td>
<td>Identical to the reduct</td>
</tr>
<tr>
<td>Protein content</td>
<td>0.75</td>
<td>Coinciding with all the present condition attributes</td>
<td>Identical to the reduct</td>
</tr>
</tbody>
</table>

The analysis of the results showed in Tables 5-6 highlights the efficiency of the our data to examine the considered problem. This affirmation is confirmed by the high level of the quality of classification, lying between 0.72 and 0.86 (in other words, at least 72% of the objects are perfectly classified or explained by means of the information present in the decision tables). Moreover, the third and fourth column of the Tables 5-6 show the reducts and core of the examined decision tables. For the analysis of the high quality milk (Table 6), all the information present in
the decision tables is necessary to explain the problem, whereas some condition attributes are superfluous in the analysis of the normal quality milk (Table 5).

Taking into account the information above mentioned on reducts and core, some of the most significant decision rules were obtained among all possible decision rules and represented in Tables 7-10.

Table 7. Decision rules obtained with TBC as decision attribute

<table>
<thead>
<tr>
<th>Rule</th>
<th>Season</th>
<th>Producer</th>
<th>Fat</th>
<th>Proteins</th>
<th>Cells</th>
<th>TBC</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>297</td>
<td>64</td>
<td>297</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>328</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>329</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>380</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>334</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>325</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>365</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>310</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>367</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>489</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>1</td>
<td>291</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>177</td>
<td>1</td>
<td>38</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>513</td>
<td>1</td>
<td>47</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1524</td>
<td>1</td>
<td>47</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1585</td>
<td>1</td>
<td>41</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1687</td>
<td>2</td>
<td>65</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1762</td>
<td>1</td>
<td>39</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2004</td>
<td>1</td>
<td>48</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3244</td>
<td>2</td>
<td>56</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>360</td>
<td>2</td>
<td>57</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>32</td>
<td>1</td>
<td>111</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>32</td>
<td>1</td>
<td>83</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1139</td>
<td>1</td>
<td>91</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1524</td>
<td>1</td>
<td>74</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td>82</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>1</td>
<td>76</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>1</td>
<td>122</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Decision rules obtained with somatic cells as decision attribute

<table>
<thead>
<tr>
<th>Rule</th>
<th>Season</th>
<th>Producer</th>
<th>Fat</th>
<th>Proteins</th>
<th>TBC</th>
<th>Cells</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>250</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>377</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>350</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>281</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>286</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>303</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>287</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>560</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
<td>45</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule</td>
<td>Season</td>
<td>Producer</td>
<td>TBC</td>
<td>Proteins</td>
<td>Cells</td>
<td>Fat</td>
<td>Support</td>
<td>Confidence</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------</td>
<td>-----</td>
<td>----------</td>
<td>-------</td>
<td>-----</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>322</td>
<td></td>
<td></td>
<td>1</td>
<td>38</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>355</td>
<td></td>
<td></td>
<td>1</td>
<td>47</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>246</td>
<td></td>
<td></td>
<td>1</td>
<td>472</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>402</td>
<td></td>
<td></td>
<td>1</td>
<td>78</td>
<td>65</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>476</td>
<td></td>
<td></td>
<td>1</td>
<td>78</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>325</td>
<td></td>
<td></td>
<td>1</td>
<td>105</td>
<td>76</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>2</td>
<td>388</td>
<td></td>
<td></td>
<td>1</td>
<td>59</td>
<td>67</td>
</tr>
<tr>
<td>8</td>
<td>177</td>
<td>2</td>
<td>37</td>
<td></td>
<td></td>
<td>1</td>
<td>29</td>
<td>73</td>
</tr>
<tr>
<td>9</td>
<td>1269</td>
<td>1</td>
<td>38</td>
<td></td>
<td></td>
<td>1</td>
<td>38</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>2121</td>
<td>2</td>
<td>39</td>
<td></td>
<td></td>
<td>1</td>
<td>38</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>1</td>
<td>117</td>
<td></td>
<td></td>
<td>1</td>
<td>78</td>
<td>65</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td></td>
<td></td>
<td>1</td>
<td>47</td>
<td>67</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>74</td>
<td></td>
<td></td>
<td>1</td>
<td>59</td>
<td>67</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>2</td>
<td>91</td>
<td></td>
<td></td>
<td>1</td>
<td>59</td>
<td>67</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>1</td>
<td>73</td>
<td></td>
<td></td>
<td>1</td>
<td>78</td>
<td>65</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>2</td>
<td>110</td>
<td></td>
<td></td>
<td>1</td>
<td>78</td>
<td>65</td>
</tr>
</tbody>
</table>

Table 9. Decision rules obtained with fat as decision attribute

<table>
<thead>
<tr>
<th>Rule</th>
<th>Season</th>
<th>Producer</th>
<th>Fat</th>
<th>TBC</th>
<th>Cells</th>
<th>Proteins</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>312</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>241</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>407</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>369</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2</td>
<td>418</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89</td>
</tr>
<tr>
<td>9</td>
<td>126</td>
<td>2</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>1364</td>
<td>1</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>1762</td>
<td>1</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>2121</td>
<td>2</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1</td>
<td>104</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>1</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66</td>
</tr>
</tbody>
</table>

Table 10. Decision rules obtained with proteins as decision attribute
In each table, the rows before and after the bold horizontal line describe respectively the decision rules describing the normal quality milk and the high quality milk. The discovered rules can be explained using an easy logical form of propositions; for example, in Table 7 the first rule can be read as: if the season in which the normal milk arrives is winter, then the TBC present in samples of the milk is in the range [1.00-116.21]. This rule is supported by 297 milk samples with a confidence level of 74%.

The analysis of the discovered rules highlights some interesting results, that can be summarized in the following points:

- Table 7 shows that the trend of the TBC in the normal quality milk samples (rules no. 1-19) is independent of the behaviour of the product variables proteins, cells and season; moreover, these rules are not able to give a good explanation of the phenomenon because their confidence level is only of 60-70% (it is a piece of information already known in the literature, since the trend of the TBC is linked to the sensibility of the elderly animals and/or with an immune system showing deficit to the infections); these reflections can be extended also for the rules explaining the TBC in the high quality milk samples (rules no. 20-27);

- Table 8 leads to the same consideration realized in the analysis of the Table 7 (i.e., the product variables and the season of the milk production are not “strongly” correlated with the somatic cells present in the normal and high quality milk samples); in this sense, a high level of somatic cells could be explained by the presence of inflammatory troubles in the animals, so that it is not correlated either to the parameters of localization of the firm or to production season of the milk;

- Table 9 examines the quantity of fat in the milk samples; this table shows some interesting results that can be summarized as follows:

  - during the Spring and Summer the fat content of the samples is on a medium-low level indicating a low milk quality; more precisely, the rules no. 1-2 describe the medium-low content of fat (<3.66%) present in the normal quality milk, while the rules no. 11 and 15 explain the content of fat (<3.61%) in the high quality milk samples;
  - the correlation between the fat content and the delivery date of the milk (season) is confirmed by the rules no. 3-4 and 13-14, where a medium-high level of the fat content is obtained in the Autumn-Winter period (the rules no.3-4 for the normal quality milk, the rules no. 13-14 for the high quality milk); these results,
in terms of correlation between fat content and kind of season, are confirmed by the statistical analyses, reported in the literature (C. Pinarelli, “L’influenza dello stress da caldo sulla produzione del latte”, in Il Latte, dicembre 2003, Vol. 28(11), pp. 36-38), conducted on other milk samples, with the difference that the rough set approach makes it possible to individualize the most significant information or independent variables by means of the evaluation of decision rules without a prior elimination of the inconsistent data and it is able to “find”, among the conditional attributes, the unique attributes that explain the phenomenon;

- the remarks previously mentioned can be extended to the analysis of the correlation between proteins and fat content; i.e., a low level of protein causes a low level of fat content (rule no. 5 for the normal quality milk and rule no. 12 for the high quality milk, while by means of the rules no. 7 and 16 it is possible to note that a high quantity of protein causes a high level of fat content);

- Table 10 analyses the relation between the considered condition attributes and proteins content (decision attribute); the analysis of the decision rules present in this table leads to the same consideration made in the Table 9 (note that in Table 9 the protein content is considered a condition attribute and fat content a decision attribute while in Table 10 the protein and fat content are considered decision and condition attributes respectively).

4 Conclusions

The results obtained outline a satisfactory sanitary situation in terms of TBC for the almost complete absence of pathogenic agents in the analysed samples, while the fluctuations found among the different farms, not statistically correlated to the business typology, seem to indicate non-optimized hygienic praxis; this study highlights how the system of breeding can strongly condition the quality of the milk of the final products obtained from it.

The proposed Rough Set approach made it possible to examine the main commodities’ characteristics of two different samples of milk (normal and high quality). According to the methodological point of view, we can affirm that the Rough Set approach:

- makes it possible to deal with qualitative and quantitative data (i.e.: seasons, providers);
- makes it possible to deal with inconsistent or ambiguous data that need not to be removed prior to the analysis;
- supplies useful elements of knowledge about the characteristics of the milk, such as relevance of attributes, information about their interaction (from quality of classification), minimal subsets of attributes (reducts)
conveying the relevant knowledge contained in the decision rules, set of
the non-reducible attributes (core);
- expresses the decision rules in the natural and understandable form “if ..., 
  then ...”;
- involves low cognitive efforts for the user (also in the case of large de-
  cision tables); more precisely, the user must provide only the decision at-
  tribute discretization;
- operates on original data without requiring statistical operators, such as
  average and standard deviation;
- involves elementary concepts and mathematical tools, without resorting
to any analytical structure.

On the other hand, the results that we obtained by means of the application of
the Rough Set approach allowed us to point out different interactions among the
considered variables; more precisely, this methodology highlighted that:
- there is a direct correlation between the season and some parameters,
such as proteins and fat content present in the milk samples; i.e., during
the warmer season (spring and summer) there is a low quantity of fat
and proteins, while during the Autumn and Winter there is a greater con-
tent of these commodity variables;
- the level of bacteria and somatic cells cannot be explained only by the
variables included in the decision table, since these parameters are corre-
lated with the hygienic-sanitary condition in which the dairy cows live.

These results could be explained by the fact that the extended exposure to high
environmental temperatures determines, in the milk cow, the activation of some
physiological mechanisms (decrease of food consumption, increase of perspiration,
etc.) that negatively reflect on the quantity and quality of the milk produced.
Therefore, those animals bred at higher temperatures take a lower quantity of
food, especially of fibrous fodder, which determines a less favourable situation for
the maintenance of high contents of fat in the milk. Even for the protein content,
the researchers agree to recognize that the cows bred in conditions of high tem-
perature provide milk with a lower percentage of protein, caused by the reduction
of food rations, with a consequent energetic deficit and a higher resort to amino
acids for the production of energy. In order to try to limit the damages caused by
thermic stress on the percentages of fat and protein, it is possible to act on the ra-
ton, offering highly digestible food that is rich in metabolizable energy, such as
high quality fodders, more appetizing food with a high level of moisture (C. Pi-
narelli, L’influenza dello stress da caldo sulla produzione del latte, in Il latte, n.
28(11), 2003, pp. 36-38).

In the final analysis, this methodology could be very efficient to also improve
the milk quality proved by the producers by the indication of the most “critical”
attributes conditioning the quality of milk; in fact, analysing the obtained decision
rules, it is possible, for example, to evaluate the relations between the same pro-
duct variables and the milk producer (coded in our decision tables). For this reason,
the methodology could be used as a tool to individualize the farmers that offer the
best milk and to encourage them to offer, in the future, milk with higher quality by
means of an economic incentive. Moreover, among the advantages that can be de-
ivered from this typology of analysis, there is also the possibility of rationalizing the payment of the milk on the grounds of quality that in Italy, according to the law N. 88 of 1988, is limited to the five product parameters considered the classifying elements of the basic quality of milk, independently of its final destination. A necessary step for the development of the system would be to apply the concept of "finalized quality", basing the evaluation of the quality on the type of transformations/use to which the milk is destined, and not letting the system of payment of the milk depend on climatic and/or nutritional variations, the proteic title of the milk depends on the correct nutritional balance between energy and protein, fibre and starches. (Boldoni G., Varisco G., Marcolini A., Comoldi M., Rabaioni C., Le caseine per la valorizzazione casearia del latte nel pagamento in base alla qualità, in Il Latte, febbraio 2003, pp. 70-76). Therefore, microbiological quality standards stipulated by this EU regulation may be used as a basis for developing a specific quality standard system for small ruminant milk in Europe.

References