Milk Quality Estimation Using a Machine Learning Techniques

Abstract

Milk is substantial commodity that is widely used around the world due to rich nutrients such as protein vitamins, and minerals that can be found in milk. However, processing milk is difficult and needing fast handling in order to maintain milk quality. In this study, milk grading data is studied to find proper machine learning model in order to identify milk grade quality based on several variables. Estimation determining of physical properties of milk is the fastest way to grading milk. There are 6 important milk features, there are pH, temperature, taste, odor, fat, turbidity and colour. Collected data then being processed and implemented in several machine machine learning algorithms, logistic regression (LR), extreme gradient boosting (XGBoost), adaptive boosting (ADABoost), random forest (RF) support vector machine (SVM), and k-nearest neighbours (KNN) to construct the classification model of milk identification. The results show that the SVM model has the best performance with the accuracy of 87% and indicate that the proposed method in this study can improve the estimation accuracy of grade milk based on physical properties, which provides a technical basis for predicting the quality of milk.

1. Introduction

Since milk is a complex biological liquid containing many physico-chemical complexes, it may be expected that some constituents will change as a result of processing. Many changes that result from processing can be determined best by chemical procedures. However, certain measurable physical changes may be more discernible and important to specific instances. Physical changes in milk or cream which may result from chemical changes in milk constituents are not clearly understood.

Milk contains more than 100 chemical ingredients such as water, fat, phospholipids, proteins, lactose, inorganic salts, and other primary compounds. The composition of milk is very complex. The mixture of lower fatty acids, acetones, acetaldehydes, carbon dioxide, and other volatile substances affects the odor of milk. Among them, sulfide is the main component of fresh milk odor. The flavor substances in milk are influenced by many factors, mainly produced by four forms, one of which is the reaction of milk fat, milk protein, and carbonic acid, etc. Triacylglycerols, fatty acids, diacylglycerides, saturated/polyunsaturated, and phospholipids in milk fat are directly related to the flavor of milk. The degradation products of protein, fat, and lactose in milk are fatty acids, sulfur-containing amino acids, thiamine, etc. The decomposition process of these substances will produce volatile compounds. Due to the different feed and growth environment of the cows from each dairy farm, the odor of milk produced is quite different. The content of milk protein and milk fat plays a significant role in milk quality evaluation. The process of degradation for milk fat and milk protein or the interaction between derivatives can affect the milk’s odor compounds. Therefore, the establishment of the milk detection model based on physical feature is considerable significance to the identification of milk quality.

- pH

The pH value for fresh raw milk is normally in the range of 6.4 to 6.8 and depends on the source of the milk.

The pH of milk accounts for the amount of lactic acid produced by microbial activity. The more lactic acid present, the higher the acidity. This would result in a change in taste and smell, making it unsuitable for human consumption.

pH is an important quality parameter in the dairy industry. The quality of raw milk, as well as essentially the finished product, must be monitored and maintained in any dairy industry, whether during packaging for human consumption or subsequent processing of other dairy products.

- Temperature

Here is some information and tips for storing and serving milk at home to keep it fresh as long as possible:

1. While at the grocery store, pick up milk last so it stays as cool as possible. Refrigerate promptly after you get home.
2. Ideally, milk should be stored in the refrigerator at 40 degrees F or below. Storing and serving milk at this temperature extends overall shelf-life and maximizes flavor.
3. Store your milk in the coldest part of the refrigerator, not in the door where it will be exposed to outside air every time someone opens it.
4. Milk is pasteurized to kill bacteria that could potentially cause health risks. Even so, it is not safe to leave pasteurized milk unrefrigerated for an extended time.

- Taste

1. The taste of milk, as this word is commonly used, is the sensation perceived when milk is taken into the mouth. The term “flavor,” as used in this paper, is a combination of the sensations of taste, perceived in the mouth, with those of smell, produced through the medium of the inner nasal passages. It has aided in precision to confine the term “taste” to those sensations which are perceived only in the mouth.

2. The primary taste of milk has been designated as the sum of all of the taste impressions coming from normal milk, and not influenced by feed or the secretion of abnormal milk.

3. The secondary taste includes that which is added to the primary taste from different sources, such as feed and products of disease. It is necessary to eliminate these secondary influences in order that primary tastes can be properly determined.

4. The chloride-lactose relation is one of the most important bases of milk taste. Milk samples with a high chloride-lactose ratio were judged less favorably than those of like origin where the chloride-lactose ratio was relatively low.

5. The primary taste of skim milk is practically equal to that of the whole milk from which it is separated.
6. By the application of dialysis, it was possible to separate the milk into two parts (dialyzate and residue) with extreme differences in taste. It was found that nearly all of the milk components producing the primary taste were present in the dialyzate, while the components remaining in the residue could be designated as free from taste.
7. By dialysis it has been further demonstrated that fat and protein substances, as well as certain difficultly dialyzable salt components, which go to make up a large percentage of the milk content, take only a subordinate part in the primary taste of the milk.
8. The dialysis of milk containing a pronounced feed taste has shown that the feed taste was found very largely in the residue, and appeared to be either not dialysable or in some way combined with milk fat or with other non-dialysable materials.

- Odor

Good quality milk should have a pleasantly sweet and clean flavor with no distinct aftertaste. Because of the perishability of milk and the nature of milk production and handling procedures, the development of off- flavors/odors is not uncommon. To prevent flavor/odor defects in milk, proper milk handling procedures from the farm to the consumer are essential. This guideline will describe the common flavor and odor defects found in milk and their potential causes. These defects may be classified according to the ABC's of off-flavors:

1. Absorbed/Transmitted
2. Bacterial/Microbial
3. Chemical/Enzymatic/Processing

- Fat

When you shop in the dairy case, the primary types of milk available are whole milk (3.25% milk fat), reduced-fat milk (2%), low-fat milk (1%) and fat-free milk, also known as skim milk. Each one packs 13 essential nutrients, including 8 grams of high-quality protein. Types of milk vary by percentage of milkfat, or the amount of fat that is in the milk by weight. These percentages are noted on the package and by the different cap colors to show the milkfat at a glance.

While the amount of milk fat does affect the number of calories and fat in each serving, all milk—from fat free to low-fat to organic and lactose free milk—remains a naturally nutrient-rich, simple and wholesome food. Understanding your choices and their differences can help you determine the best type of milk for each member of your family.

- Turbidity

The turbidity per unit concentration of total solids varies with the fat content of the milk and also with the efficiency of homogenization since about three- fourths of the total turbidity can be attributed to the fat phase. When the turbidity due to the non-fat portion of the milk has been reduced to a negligible quantity by the addition of ammonia, the turbidity of the fat globules per unit concentration of fat is shown to be closely related to the particle size distribution.

- Colour

Milk is a natural, whole food made up of water, protein, fat, carbohydrates in the form of lactose, vitamins including calcium, minerals including phosphorous and a range of other bioactive compounds.

Caseins are one of the main types of protein in milk which cluster together with calcium and phosphate to form tiny particles called micelles. When light hits these casein micelles it causes the light to refract and scatter resulting in milk appearing white.1

Previously, when milk was delivered in bottles with aluminium tops, the yellow coloured fat or cream of cow's milk would separate and rise to the top of the bottle producing a pale coloured milk. Today, most milks are homogenised which passes the milk under pressure through very fine nozzles, evenly dispersing the fat and protein micelles to create a smooth, creamy texture and taste, plus brighter white colour.

The Australian Food Standards Code allows the components of milk, such as lactose, protein, fat or vitamins and minerals to be adjusted by adding or removing those components to produce a consistent product. In earlier times natural variations in colour that would have occurred in milk because of differences between cow breeds or the pasture are now standardised and the bright, white colour consistent. There are no artificial colours added to give milk its white appearance in Australia.

Therefore, this study proposes a fast identification method based on physical properties and machine learning techniques for milk milk quality estimation. The collected data are preprocessed, pattern recognition algorithms are used for modeling to achieve the detection target. Based on six different classification algorithms: logic regression (LR), support vector machine (SVM), and random forest (RF), extreme gradient boosting (XGBoost), adaptive boosting (ADABoost) and k-nearest neighbours (KNN) are used to construct models to estimate the milk quality.

2.3.3. LR

Logistic regression is a supervised machine learning algorithm for solving classification problems. The principle is to find the minimum value of the loss function to make the prediction function more accurate, thereby solving the classification problem. The penalty term is a vital hyperparameter of the LR model, and the solver parameter can optimize the loss function [34].

2.3.5. XGBoost

Extreme gradient boosting algorithm is an improved version based on GBDT, which is not sensitive to input requirements and is widely used in the industry. Compared with the general

GBDT algorithm, XGBoost uses the second derivative of the loss function about the function to be sought, adds a regularisation term to prevent overfitting, and samples the attributes when constructing each tree. It has fast training speed and high accuracy and fitting effect, etc. [36].

2.3.5. ADABoost

AdaBoost stands for Adaptive Boosting. Literally, boosting means to arrange a set of weak classifiers in a sequence in which each weak classifier is the best choice for a classifier at that point for rectifying the errors made by the previous classifier.] In the sequence of weak classifiers used, each classifier focuses its discriminatory firepower at the training samples misclassified by the previous weak classifier.

2.3.2. RF

Random forest is a crucial bagging-based ensemble learning method. It is composed of many decision trees (CARTs). It can be used to solve classification and regression problems and has strong anti-noise ability, can avoid overfitting. The procedure of developing an RF model is as follows: firstly, m sample points are extracted from the training sample set S to form a new training subset; secondly, a classification decision tree or regression model is constructed for each training subset, which is obtained by randomly selecting k features among all features as split nodes; the output of the model is the category (classification) with the highest number of votes or the average output (regression) of each decision tree [33].

2.3.1. SVM

SVM is a supervised learning model that can perform pattern recognition, classification, and regression analysis [31,32]. The principle of SVM is to find the separation hyperplane that can correctly divide the classes in the training set and obtain the largest geometric distance. The objective function of the SVM is as follows:

$$minw,b,\xi 12wTw+C\sum i=1n\xi is.t.yi(wT\phi(xi)+b)\geq 1-\xi i,\xi i\geq 0,i=1,2,\ldots,n$$

(2)

where w and b are the SVM parameters, ξi is the classification loss of the *i*th sample point, $\phi(xi)$ is the mapping function, C is the penalty parameter, x is the *i*th input sample, and n is the number of training samples.

For nonlinear classification problems, the kernel (mapping) function in SVM can map samples from the original space to high-dimensional space, making the samples linearly separable in the new space. Among them, the most commonly used and the most effective is the radial basis kernel function (RBF kernel):

$$K(x1,x2)=\exp(-\gamma\|x1-x2\|2),\gamma>0$$

(3)

where x1, x2 are sample points of traing set; the parameter γ (gamma), defines the range of influence for a single training example, with low values meaning ‘far’ and high values meaning ‘close’.

2.3.4. KNN

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.

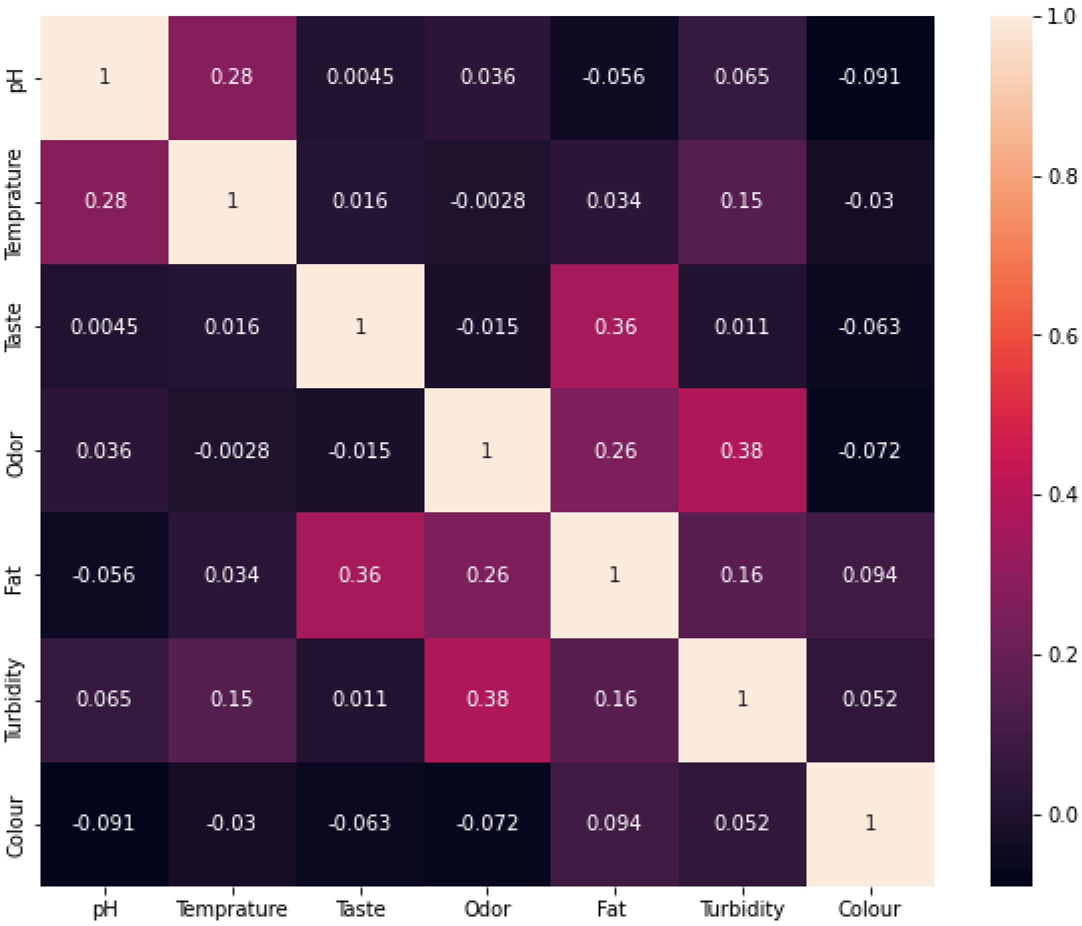
2. Materials and Methods

In this study, dataset is obtained from Kaggle and processing data is done. Then duplicate data are handled by deleting it in order to better model, after that non integer data are labeled. Due to unbalanced data value for each grade, oversampling is implemented. Furthermore, x and y variables are defined and training and testing data are separated. Lastly, all of value of x variables are standardized.

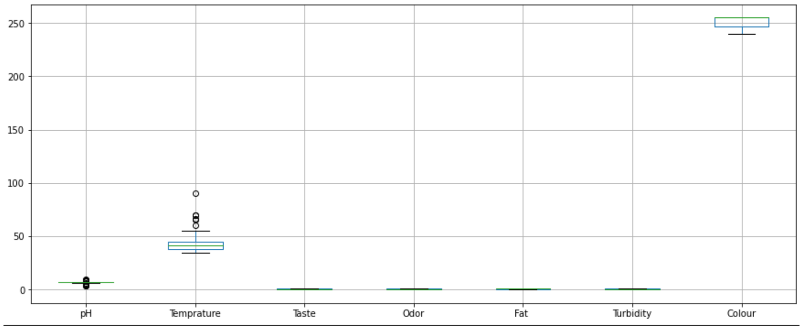
Processing data is important to fit values in machine learning model. Six machine learning algorithms, logistic regression (LR), extreme gradient boosting (XGBoost), adaptive boosting (ADABoost), random forest (RF) support vector machine (SVM), and k-nearest neighbours (KNN), are then used to construct the classification models.

3. Results

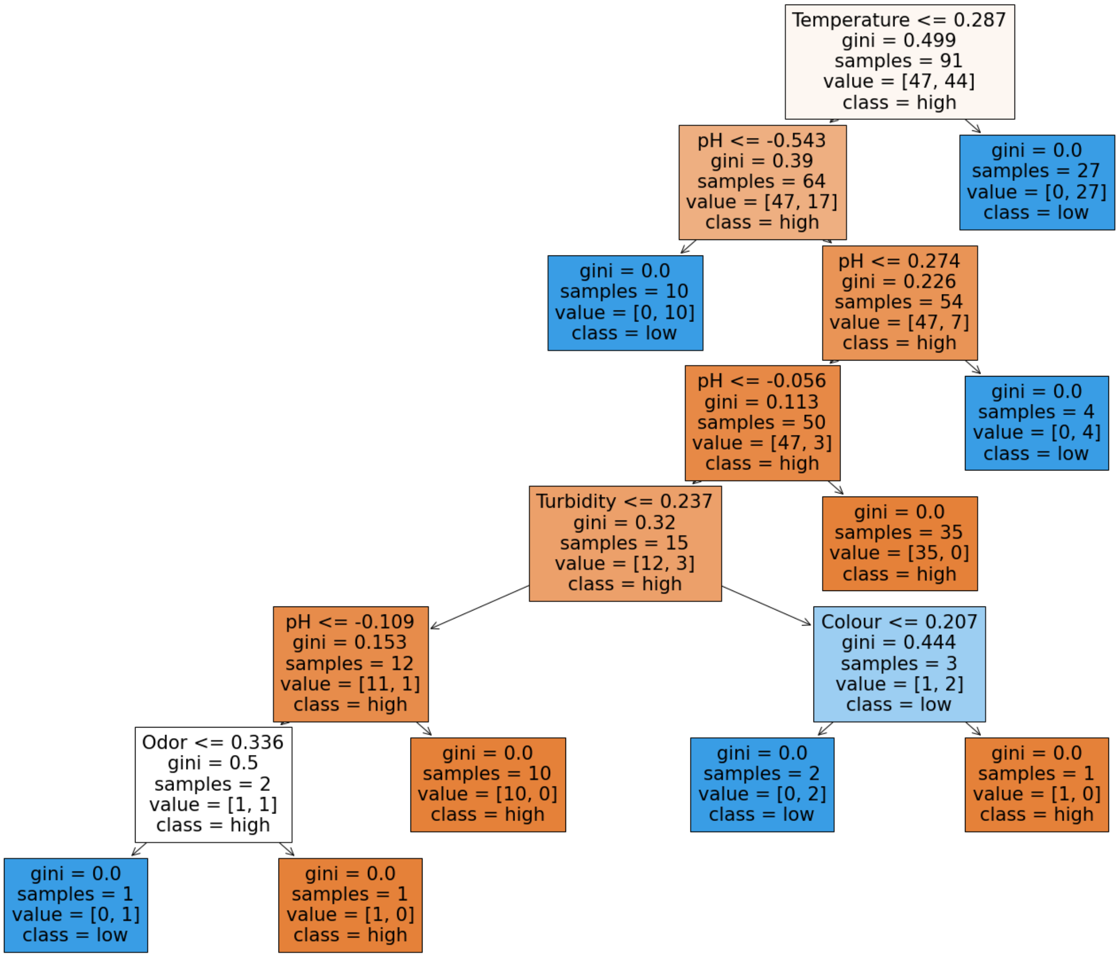
Understanding data is the first thing can be done. Checking dimension, null value, data value, data shape, etc are the most important process to understand data. Initially, dataset has 1059 x 8 dimension. Then, duplicated data is being processed. After checking the skewness, pH, temperature, fat and colour column > +- 0.5, imputated by median and the rest with mean.



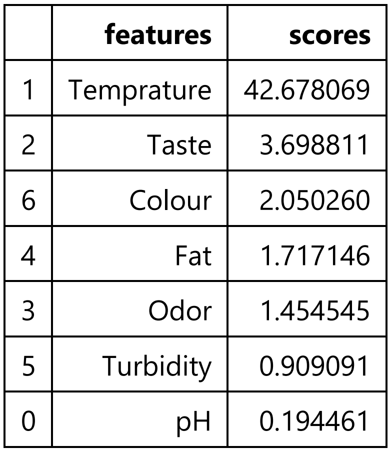
Pertaining to the pearson correlation calculation, turbidity and odor has correlation better than other variables with correlation value of 0.38.



Imbalance data for every variables also checked and it is found that the data is reasonably good. In order to fit the machine learning algorithm, data should be integer. Hence grade variables encoded into integer where 0 stands for high grade and 1 for low grade. Finally, every variables value are standardized to gain better model and calculation.



Logistic tree regression model algorithm is implemented to the processed data and showing the deep of the tree. However, the accuracy of test data from using this model is still low with only 76%. Cross validation and feature analysis are also implemented to the model result.



	features	scores
1	Temprature	42.678069
2	Taste	3.698811
6	Colour	2.050260
4	Fat	1.717146
3	Odor	1.454545
5	Turbidity	0.909091
0	pH	0.194461

The result is that Temperature, Odor, Fat, Taste, and Colour have more impact to the grade. So after implementing cross validation and hyperparameter analysis, the accuracy rise in to 83%. Furthermore, other 5 alghorithm which are extreme gradient boosting (XGBoost), adaptive boosting (ADABoost), random forest (RF) support vector machine (SVM), and k-nearest neighbours (KNN), are then used to implemented to the processed data.

	mae	mse	mape	acc	r2
Decision Tree	0.000000	0.000000	0.000000e+00	0.739130	-0.045455
XGBoost	0.000000	0.000000	0.000000e+00	0.782609	0.115385
Random Forest	0.000000	0.000000	0.000000e+00	0.826087	0.269841
ADABoost	0.000000	0.000000	0.000000e+00	0.739130	-0.045455
SVM	0.252747	0.252747	8.908219e+14	0.869565	0.469231
KNN	0.098901	0.098901	3.959208e+14	0.739130	-0.095238

The result is support vector machine gives the highest accuracy for data test with accuracy value of 87%. This result is also supported by Fanglin Mu, Yu Gu, Jie Zhang, and Lei Zhang, 2020 study that SVM algorithm is suitable for grading milk quality.

Conclusion

Milk is substantial commodity that is widely used around the world due to rich nutrients such as protein vitamins, and minerals that can be found in milk. However, processing milk is difficult and needing fast handling in order to maintain milk quality. The results show that the SVM model has the best performance with the accuracy of 87% and indicate that the proposed method in this study can improve the estimation accuracy of grade milk based on physical properties, which provides a technical basis for predicting the quality of milk.

References

1. Fanglin Mu, Yu Gu, Jie Zhang, and Lei Zhang. 2020. Milk Source Identification and Milk Quality Estimation Using an Electronic Nose and Machine Learning Techniques. National Library of Medicine. Doi:10.3390/s20154238

2. Zhang J., Yang M., Cai D., Hao Y., Zhao X., Zhu Y., Zhu H., Yang Z. Composition, coagulation characteristics, and cheese making capacity of yak milk. *J. Food Sci.* 2020;**103**:1276–1288. doi: 10.3168/jds.2019-17231. [PubMed] [CrossRef] [Google Scholar]

3. Bilandzic N., Dokic M., Sedak M., Solomun B., Varenina I., Knezevic Z., Benic M. Trace element levels in raw milk from northern and southern regions of Croatia. *Food Chem.* 2011;**127**:63–66. doi: 10.1016/j.foodchem.2010.12.084. [CrossRef] [Google Scholar]

4. Tamsma A., Kurtz F.E., Bright R.S., Pallansch M.J. Contribution of milk fat to the flavor of milk. *J. Dairy Sci.* 1969;**52**:1910–1913. doi: 10.3168/jds.S0022-0302(69)86872-4. [CrossRef] [Google Scholar]

5. Kinsella J.E., Patton S., Dimick P.S. The flavor potential of milk fat. A review of its chemical nature and biochemical origin. *J. Am. Oil Chem. Soc.* 1967;**44**:449–454. doi: 10.1007/BF02666792. [CrossRef] [Google Scholar]

6. Forss D.A. Mechanisms of formation of aroma compounds in milk and milk products. *J. Dairy Res.* 1979;**46**:691–706. doi: 10.1017/S0022029900020768. [CrossRef] [Google Scholar]

7. Mcgorrin R.J. Flavor analysis of dairy products. *ACS Symp. Ser.* 2007;**971**:23–49. doi: 10.1021/bk-2007-0971.ch002. [CrossRef] [Google Scholar]

7. Keenan T.W., Lindsay R.C. Evidence for a dimethyl sulfide precursor in milk. *J. Dairy Sci.* 1968;**51**:112–114. doi: 10.3168/jds.S0022-0302(68)86929-2. [CrossRef] [Google Scholar]

8. Faulkner H., O'Callaghan T.F., McAuliffe S., Hennessy D., Stanton C., O'Sullivan M.G., Kerry J.P., Kilcawley K.N. Effect of different forage types on the volatile and sensory properties of bovine milk. *J. Dairy Sci.* 2018;**101**:1034–1047. doi: 10.3168/jds.2017-13141. [PubMed] [CrossRef] [Google Scholar]

9. Kuhn J., Considine T., Singh H. Interactions of milk proteins and volatile flavor compounds: Implications in the development of protein foods. *J. Food Sci.* 2006;**71**:72–82. doi: 10.1111/j.1750-3841.2006.00051.x. [CrossRef] [Google Scholar]

10. https://www.journalofdairyscience.org/article/S0022-0302(29)93594-3/fulltext

11. https://www.usdairy.com/news-articles/how-long-can-milk-sit-out

12. https://www.mt.com/hk/en/home/library/applications/lab-analytical-instruments/measurement-pH-of-milk.html

13. https://gonnaneedmilk.com/articles/types-of-milk-explained/

14. https://www.dairy.com.au/dairy-matters/you-ask-we-answer/why-is-milk-white