A COMPARISON OF MODELS ON THE PREDICTION OF MILK QUALITY

As an avid milk drinker, quality is everything for the product. I have chosen to conduct a comparative and analytical study on milk quality prediction, using multiple machine learning methods. In this report, I detail the findings of three scholarly sources as well as my own findings on the matter.

SCHOLARLY ARTICLES

The results in Table A come from the American Dairy Science Association(3) and show the use of multiple models on the prediction of milk quality. Of these models I will be comparing the Support Vector Machine(SVM) and the Logistic Regression. The models reach and accuracy of .73 and .75 respectively.

The next article from the Hebei University of Technology(2) uses a Random Forrest model, a SVM model and a Logistic Regression model. The results show accuracy scores of .94, .95, and .925 respectively. These results can be seen in Table B.

The last article in my comparison comes from the Eurasian Journal of Food Science and Technology(1). The results from this paper can be seen in Table C. The journal use and Neural Network as well as AdaBoost to predict on milk quality. The Neural Net which had an Accuracy score of .997, an F1 score of .955, a Precision score of .959, and a Recall of .954. The AdaBoost model had a score of .99 for each metric.

	Best classification	
Trait	Method[1]	Accuracy[2] (SD)
Technological[3]		
RCT	PLSDA, SVM	0.75 (0.03, 0.06)
k20	SVM	0.73 (0.02)
a30	SVM	0.73 (0.03)
a60	PLSDA	0.69 (0.07)
CN micelle size	SVM	0.62 (0.03)
pH	PLSDA, SVM	0.80 (0.03, 0.02)
Heat stability	PLSDA, SVM	0.74 (0.04, 0.05)
Protein		
α_{S1}-CN	RF	0.48 (0.02)
α_{S2}-CN	PLSDA	0.40 (0.04)
β-CN	SVM	0.46 (0.04)
κ-CN	RF	0.45 (0.02)
α-LA	SVM	0.43 (0.03)
β-LG A	PLSDA	0.42 (0.03)
β-LG B	PLSDA	0.41 (0.04)

(a) Dairy Science Results (3)

Features		SVM		RF		LR	
		Train	Test	Train	Test	Train	Test
DHI	PCA	19.50	15.50	17.63	18.50	19.88	18.00
	LDA	57.75	58.50	52.13	53.50	53.38	56.00
E-nose	PCA	56.25	59.50	71.62	70.50	62.00	65.00
	LDA	85.75	85.00	82.13	80.50	84.38	81.50
Fusion	PCA	41.50	45.00	53.38	51.50	39.75	34.50
	LDA	95.50	95.00	92.50	94.00	93.50	92.50

(b) MDIP Results (2)

Model Name	AUC	CA	F1	Precision	Recall
AdaBoost	0.999	0.999	0.999	0.999	0.999
Neural Network	0.997	0.954	0.955	0.959	0.954

(c) Eurasian Milk Results (1)

My Results

During my analysis, I chose eight machine learning models to test and train on the milk quality dataset. These models included Logistic Regression, Neural Network, SVM, Decision Tree, Random Forrest(RF), AdaBoost, k-Nearest Neighbors(kNN), and Gradient Boosting(GBM). Of these models, my Random Forrest, Neural Net, and SVM, all outperformed those in the scholarly articles mentioned above. However, my AdaBoost as well as my Logistic Regression fell below the estimates found in the scholarly articles.

model	accuracy	fScore	precision	recall
Log Reg	0.849	0.842	0.84	0.849
Neural Network	0.997	0.996	0.996	0.997
SVM	0.994	0.993	0.993	0.993
Tree	0.991	0.989	0.99	0.988
RF	0.994	0.994	0.993	0.994
AdaBoost	0.915	0.9	0.921	0.89
kNN	0.997	0.996	0.996	0.997
GBM	0.994	0.993	0.993	0.993

Figure 0.1: My Results

My AdaBoost received an accuracy score of .915 compared to the other model which had a score of .99(1). My Logistic Regression model received and accuracy score of .849 compared to the other model which received a score of .925(2). The reason that my models performed worse than those in the scholarly articles is most likely due to outliers. During my outlier removal process, I found almost half the data being removed. I was not willing to suffer that data loss and felt that it would damage the validity of the data. I believe that my AdaBoost and Logistic Regression models suffer as result.

REFERENCES

[1] Çelik, A. (2022). *Using Machine Learning Algorithms to Detect Milk Quality*. Eurasian Journal of Food Science and Technology, 6(2), 76–87. https://dergipark.org.tr/en/download/article-file/2673216.

[2] Mu, F., Gu, Y., Zhang, J., & Zhang, L. (2020). *Milk Source Identification and Milk Quality Estimation Using an Electronic Nose and Machine Learning Techniques*. Sensors, 20(15), 4238. https://doi.org/10.3390/s20154238.

[3] Frizzarin, M., Gormley, I. C., Berry, D. P., Murphy, T. B., Casa, A., Lynch, A., & McParland, S. (2021). *Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods*. Journal of Dairy Science, 104(7), 7438–7447. https://doi.org/10.3168/jds.2020-19576.