

Table 1: RMSE and R^2 values of 2 Regression algorithms (GLM and LM) and 2 Classification algorithms (RF and RPART) trained to predict CH_4 emission yield based on Management (Cow ID and DMI), and TSS normalised relative abundance of several bacteria taxa (Proteobacteria, Plantomycetota, Bacteoidota, Cyanobacteria, Firmicutes C). There was no significant difference in RMSE between the two types of algorithms ($P=0.97$) but the R^2 was significantly higher for Classification methods ($P=0.01$).

Regression Methods				Classification Methods		
CV	Method	RMSE	R^2	Method	RMSE	Best R^{2a}
20-Fold	glm ¹	3.70	0.19	rpart ³	3.55	0.72
15-Fold	glm	3.78	0.17	rpart	3.56	0.31
10-Fold	glm	3.23	0.31	rpart	3.43	0.19
5-fold	glm	3.62	0.21	rpart	4.33	0.25
20-Fold	lm ²	3.80	0.22	rf ⁴	2.85	0.64
15-Fold	lm	3.52	0.22	rf	2.99	0.64
10-Fold	lm	3.54	0.22	rf	3.05	0.66
5-fold	lm	3.68	0.24	rf	3.06	0.78

1. Generalised linear model
2. Linear Regression Model
3. Decision Trees
4. Random Forests

a. $P=0.01$