

Predicting Foodborne Disease Outbreaks with Food Safety Certifications:

An Machine Learning Approach for the United States and Europe

Yuqing Zheng (University of Kentucky, US), Azucena Gracia (CITA, Spain), Lijiao Hu (California State University, Chico, US)

Background & Aims

In 2020, 299 and 3,166 foodborne disease outbreaks occurred in the U.S. and Europe, posing a significant global health burden.

Since the late 1990s, food safety certification has emerged as a prominent and influential regulatory mechanism in both the private and public spheres of the agri-food system.

The major standards are: GlobalG.A.P., British Retail Consortium (BRC), FSSC, IFS, ISO 22000, PrimusGFS, and Safe Quality Food (SQF), Hu et al. (2023).

Objectives:

- Investigate the association between foodborne disease outbreaks and the adoption of food safety certifications.
- Use machine learning techniques to examine how well we can use food safety certification data to predict foodborne disease outbreaks.

Model

U.S.: State Foodborne illness = F(certifications to SQF, PrimusGFS, BRC, USDAGAP, GlobalGAP, controls)

Europe: Country Foodborne illness = G(certifications to GlobalGAP, ISO 22000, FSSC, controls)

Data sources: U.S. Centers for Disease Control and Prevention, European Food Safety Authority Dashboard.

Table 1. Regression Results

	(1) U.S. 2016, 2018–2020	(2) Europe, 2017–2020
SQF	-1.989* (1.02)	
GFS	-0.099** (0.04)	
BRC	-5.580* (3.19)	
Gap	0.57 (0.53)	
GlobalG.A.P.	0.175 (0.21)	0.099 (0.23)
FSSC22000	-4.974* (2.92)	-2.527* 1.38
ISO22000		-0.780* (0.46)
GDP	1.748** (0.73)	-1.551*** (0.53)
R ²	0.804	0.659
N	173	119

* p<0.10, ** p<0.05, *** p<0.01.

Machine Learning

- Three algorithms: OLS, decision tree, and random forest.
- Tree can capture non-linear relationships.
- Random forest combines the results of multiple trees and automatically detects interactions to improve prediction.

Table 2. Machine Learning Results for the United States

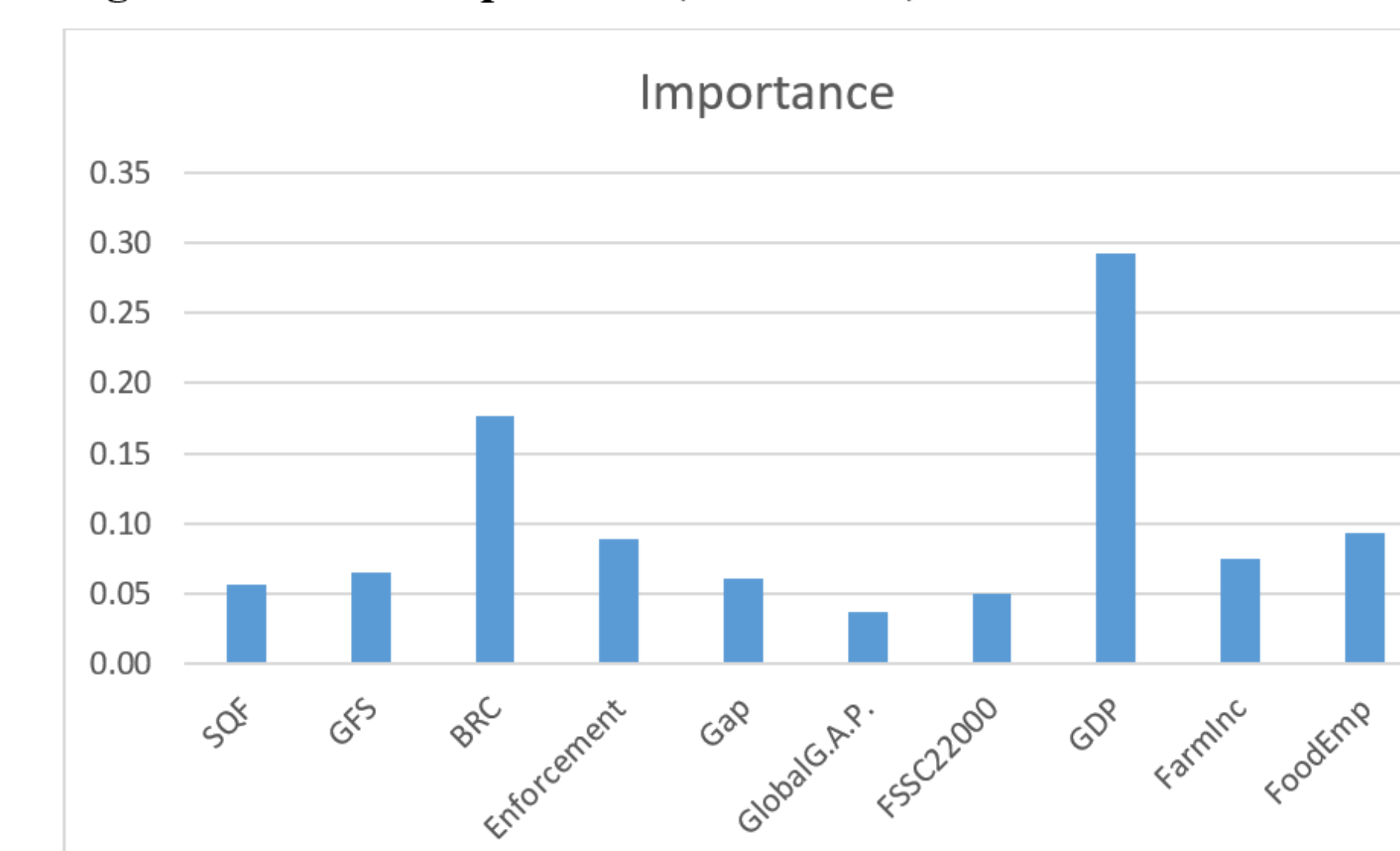
a. Predicting the number of illnesses

Methods	Training Accuracy	Testing Accuracy	Standard Errors (Testing Accuracy)
OLS	0.88	0.73	0.20
Decision Tree	0.99	0.74	0.36
Random Forest	0.97	0.76	0.31

b. Predicting the number of deaths

Methods	Training Accuracy	Testing Accuracy	Standard Errors (Testing Accuracy)	Classification Error Rate	
				Training	Testing
Multinomial	0.89	0.81	0.14	0.12	0.08
Decision Tree	0.92	0.81	0.12	0.09	0.19
Random Forest	1.00	0.80	0.10	0.00	0.17

Figure 1. Feature Importance (U.S. Model)



Conclusion

- A negative association between food safety certification association and foodborne disease outbreaks
- U.S.: Certifications to SQF, PrimusGFS, BRC, or FSSC
- Europe: Certifications to ISO 22000 or FSSC.
- Through machine learning, our models with food safety certification adoption can predict the U.S. state-level foodborne illnesses with a relatively high degree of precision (testing accuracy at around 75%).
- Certification alone could be the second most important variable (after GDP) explaining foodborne disease outbreaks.

References

Hu, Lijiao, Yuqing Zheng, Timothy A. Woods, Yoko Kusunose, and Steven Buck. "The market for private food safety certifications: Conceptual framework, review, and future research directions." *Applied Economic Perspectives and Policy* 45, no. 1 (2023): 197-220.