



This is a repository copy of *Ensemble decision tree models using RUSBoost for estimating risk of iron failure in drinking water distribution systems.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/112511/>

Version: Supplemental Material

Article:

Mounce, S.R. orcid.org/0000-0003-0742-0908, Ellis, K., Edwards, J.M. et al. (3 more authors) (2017) Ensemble decision tree models using RUSBoost for estimating risk of iron failure in drinking water distribution systems. *Water Resources Management*, 31 (5). pp. 1575-1589. ISSN 0920-4741

<https://doi.org/10.1007/s11269-017-1595-8>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Ensemble decision tree models using RUSBoost for estimating risk of iron failure in drinking water

distribution systems

Supplementary Material 1 – all matrix parameters

Data Matrix Parameter List

Parameter name	Definition
dma	DMA reference number
wqz	WQZ reference number
population	Number of people supplied by DMA
iron_Nsamples	Number of samples featuring an iron measurement in the DMA, over the year
iron_Nfails	Number of iron failures in the DMA, over the year
iron_av	Median average iron content (mg/l) in the DMA, over the year
mang_Nsamples	Number of samples featuring a manganese measurement in the DMA, over the year
mang_Nfails	Number of manganese failures in the DMA, over the year
mang_av	Median average manganese content (mg/l) in the DMA, over the year
turb_Nsamples	Number of samples featuring a turbidity measurement in the DMA, over the year
turb_Nfails	Number of turbidity failures in the DMA, over the year
turb_av	Median average turbidity (FTU) in the DMA, over the year
chlor_total_Nsamples	Number of samples featuring a chlorine measurement in the DMA, over the year
chlor_total_av	Median average total chlorine content (mg/l) in the DMA, over the year
chlor_free_Nsamples	Number of samples featuring a chlorine measurement in the DMA, over the year
chlor_free_av	Median average total chlorine content (mg/l) in the DMA, over the year
pH_Nsamples	Number of samples featuring a pH measurement in the DMA, over the year
pH_av	Median average pH content (mg/l) in the DMA, over the year
temp_Nsamples	Number of samples featuring a temperature measurement in the DMA, over the year
temp_av	Median average temperature (°C) in the DMA, over the year
cc	Number of discolouration customer contacts in the DMA, over the year
cc_clusters	Number of discolouration customer contacts in the DMA that are part of WQZ-level clusters, over the year
iron_lined	Number of km of lined cast iron pipe main in DMA
iron_unlined	Number of km of unlined cast iron pipe in DMA
other_material	Number of km of non-iron pipe main (lined or unlined) in DMA

Ensemble decision tree models using RUSBoost for estimating risk of iron failure in drinking water distribution systems

Parameter name	Definition
total_length	Total length of pipe main in DMA
wtw_iron_Nsamples	Number of samples used for calculating median average iron content (mg/l) in supply to the DMA from WTW(s) over the year
wtw_iron_av	Median average iron content (mg/l) in supply to the DMA from WTW(s) over the year
wtw_mang_Nsamples	Number of samples used for calculating median average manganese content (mg/l) in supply to the DMA from WTW(s), over the year
wtw_mang_av	Median average manganese content (mg/l) in supply to the DMA from WTW(s), over the year
wtw_turb_Nsamples	Number of samples used for calculating median average turbidity (FTU) in supply to the DMA from WTW(s), over the year
wtw_turb_av	Median average turbidity (FTU) in supply to the DMA from WTW(s), over the year
wtw_chlor_total_Nsamples	Number of samples used for calculating median average total chlorine content (mg/l) in supply to the DMA from WTW(s), over the year
wtw_chlor_total_av	Median average total chlorine content (mg/l) in supply to the DMA from WTW(s), over the year
wtw_chlor_free_Nsamples	Number of samples used for calculating median average free chlorine content (mg/l) in supply to the DMA from WTW(s), over the year
wtw_chlor_free_av	Median average free chlorine content (mg/l) in supply to the DMA from WTW(s), over the year
wtw_ph_Nsamples	Number of samples used for calculating median average pH of water supplied to the DMA from WTW(s), over the year
wtw_ph_av	Median average pH of water supplied to the DMA from WTW(s), over the year
wtw_temp_Nsamples	Number of samples used for calculating median average temperature of water supplied to the DMA from WTW(s), over the year
wtw_temp_av	Median average temperature of water supplied to the DMA from WTW(s), over the year
srv_iron_Nsamples	Number of samples used for calculating median average iron content (mg/l) in supply to the DMA from SRV(s) over the year
srv_iron_av	Median average iron content (mg/l) in supply to the DMA from SRV(s) over the year
srv_mang_Nsamples	Number of samples used for calculating median average manganese content (mg/l) in supply to the DMA from SRV(s), over the year

Ensemble decision tree models using RUSBoost for estimating risk of iron failure in drinking water distribution systems

Parameter name	Definition
srv_mang_av	Median average manganese content (mg/l) in supply to the DMA from SRV(s), over the year
srv_turb_Nsamples	Number of samples used for calculating median average turbidity (FTU) in supply to the DMA from SRV(s), over the year
srv_turb_av	Median average turbidity (FTU) in supply to the DMA from SRV(s), over the year
srv_chlor_total_Nsamples	Number of samples used for calculating median average total chlorine content (mg/l) in supply to the DMA from SRV(s), over the year
srv_chlor_total_av	Median average total chlorine content (mg/l) in supply to the DMA from SRV(s), over the year
srv_chlor_free_Nsamples	Number of samples used for calculating median average free chlorine content (mg/l) in supply to the DMA from SRV(s), over the year
srv_chlor_free_av	Median average free chlorine content (mg/l) in supply to the DMA from SRV(s), over the year
srv_pH_Nsamples	Number of samples used for calculating median average pH of water supplied to the DMA from SRV(s), over the year
srv_pH_av	Median average pH of water supplied to the DMA from SRV(s), over the year
srv_temp_Nsamples	Number of samples used for calculating median average temperature of water supplied to the DMA from SRV(s), over the year
srv_temp_av	Median average temperature of water supplied to the DMA from SRV(s), over the year
Nflushes_routine	Number of routine flushing operations in DMA, over the year
Nflushes_reactive	Number of reactive flushing operations in DMA, over the year
Nbursts	Number of bursts in DMA, over the year