Thème : Projet de visualisation   TP 4

Description and relevant informations

a) Title: cylinder bands; name of the file: bands3.csv

b) Abstract:

Machine learning tools show significant promise for knowledge acquisition, particularly when human expertise is inadequate. Recently, process delays known as cylinder banding in rotogravure printing were substantially mitigated using control rules discovered. Our work exemplifies a more general methodology which transforms the knowledge acquisition task from one in which rules are directly elicited from an expert, to one in which a learning system is responsible for rule generation. The primary responsibilities of the human expert are to evaluate the merits of generated rules, and to guide the acquisition and classification of data necessary for machine induction. These responsibilities require the expert to do what an expert does best: to exercise his or her expertise. This seems a more natural fit to an expert's capabilities than the requirements of traditional methodologies that experts explicitly enumerate the rules that they employ.

c) Number of instances: 512

d) Number of attributes: 40 including the class attribute – 20 attributes are numeric, 20 are nominal

e) Attribute Information:

1. timestamp: numeric; $19500101 - 21001231$
2. cylinder number: nominal
3. customer: nominal
4. job number: nominal
5. grain screened: nominal; yes, no
6. ink color: nominal; key, type
7. proof on ctd ink: nominal; yes, no
8. blade mfg: nominal; benton, daetwyler, uddeholm
9. cylinder division: nominal; gallatin, warsaw, mattoon
10. paper type: nominal; uncoated, coated, super
11. ink type: nominal; uncoated, coated, cover
12. direct steam: nominal; use; yes, no
13. solvent type: nominal; xylol, lactol, naptha, line, other
14. type on cylinder: nominal; yes, no
15. press type: nominal; use; 70 wood hoe, 70 motter, 70 albert, 94 motter
16. press: nominal; 821, 802, 813, 824, 815, 816, 827, 828
17. unit number: nominal; 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
18. cylinder size: nominal; catalog, spiegel, tabloid
19. paper mill location: nominal; north us, south us, canadian, scandanavian, mid european

20. plating tank: nominal; 1910, 1911, other

21. proof cut: numeric; $0 - 100$

22. viscosity: numeric; $0 - 100$

23. caliper: numeric; $0 - 1.0$

24. ink temperature: numeric; $5 - 30$

25. humifity: numeric; $5 - 120$

26. roughness: numeric; $0 - 2$

27. blade pressure: numeric; $10 - 75$

28. varnish pct: numeric; $0 - 100$

29. press speed: numeric; $0 - 4000$

30. ink pct: numeric; $0 - 100$

31. solvent pct: numeric; $0 - 100$

32. ESA Voltage: numeric; $0 - 16$

33. ESA Amperage: numeric; $0 - 10$

34. wax: numeric ; $0 - 4.0$

35. hardener: numeric; $0 - 3.0$

36. roller durometer: numeric; $15 - 120$

37. current density: numeric; $20 - 50$

38. anode space ratio: numeric; $70 - 130$

39. chrome content: numeric; $80 - 120$

40. band type: nominal; class; band, no band

f) Missing attribute values : yes, in 302 examples

g) Class distribution : (out of 512 total instances)

  1. 312 no band
  2. 200 band

h) **Goal** : you must use the CRISP-DM model to identify the root causes of banding.