DATA SET DESCRIPTION The Avila data set has been extracted from 800 images of the the "Avila Bible", a giant Latin copy of the whole Bible produced during the XII century between Italy and Spain. The palaeographic analysis of the manuscript has individuated the presence of 12 copyists. The pages written by each copyist are not equally numerous. Each pattern contains 10 features and corresponds to a group of 4 consecutive rows. The prediction task consists in associating each pattern to one of the 12 copyists (labeled as: A, B, C, D, E, F, G, H, I, W, X, Y). The data have has been normalized, by using the Z-normalization method, and divided in two data sets: a training set containing 10430 samples, and a test set containing the 10437 samples. Class distribution (training set) A: 4286 B: 5 C: 103 D: 352 E: 1095 F: 1961 G: 446 H: 519 I: 831 W: 44 X: 522 Y: 266 ATTRIBUTE DESCRIPTION ID Name F1 intercolumnar distance F2 upper margin F3 lower margin F4 exploitation F5 row number F6 modular ratio F7 interlinear spacing F8 weight F9 peak number F10 modular ratio/ interlinear spacing Class: A, B, C, D, E, F, G, H, I, W, X, Y CITATIONS If you want to refer to the Avila data set in a publication, please cite the following paper: C. De Stefano, M. Maniaci, F. Fontanella, A. Scotto di Freca, Reliable writer identification in medieval manuscripts through page layout features: The "Avila" Bible case, Engineering Applications of Artificial Intelligence, Volume 72, 2018, pp. 99-110.