AlvsPK Challenge: FACT SHEET

Title: Boosting with SVM
Name, address, email: Jorge Sueiras, jorge.sueiras@neo-metrics.com
Acronym of your best entry: Boost tree

Method:
For NOVA
1.- Parsing each record, generating a word count table. Usage of a thesaurus and a word exclusion list.
2.- Application of SVM techniques for dimensional reduction.
3.- Selection of input variables: Top 50 variables from the SVM process and top 100 variables from word counting.
4.- Model building through application of bootstrapping techniques to decision tree models with no more than 8 leaves per tree. Tree splitting search criterion was based on the Chi-square test.

Results:

Table 1: Our methods best results

Dataset	Entry name	Entry ID	Test BER	Test AUC	Score	Track
ADA	boost tree	906	0.1825	0.8075	0.1575	Prior
GINA	Boost mix	915	0.0896	0.9655	0.6581	Prior
HIVA	boost tree	906	0.3257	0.7127	0.739	Agnos
NOVA	boost tree	906	0.0507	0.9869	0.2115	Agnos
SYLVA	boost tree	906	0.0081	0.9996	0.2211	Prior
Overall	boost tree	906	0.1314	0.8944	0.3983	Prior

Table 2: Winning entries of the AlvsPK challenge

Best results agnostic learning track						
Dataset	**Entrant name**	**Entry name**	**Entry ID**	**Test BER**	**Test AUC**	**Score**
ADA	Roman Lutz	LogitBoost with trees	13, 18	0.166	0.9168	0.002
GINA	Roman Lutz	LogitBoost/Doubleboost	892, 893	0.0339	0.9668	0.2308
HIVA	Vojtech Franc	RBF SVM	734, 933, 934	0.2827	0.7707	0.0763
NOVA	Mehreen Saeed	Submit E final	1038	0.0456	0.9552	0.0385
SYLVA	Roman Lutz	LogitBoost with trees	892	0.0062	0.9938	0.0302
Overall	Roman Lutz	LogitBoost with trees	892	0.1117	0.8892	0.1431
Best results prior knowledge track						
Dataset	**Entrant name**	**Entry name**	**Entry ID**	**Test BER**	**Test AUC**	**Score**
ADA	Marc Boulle	Data Grid	920, 921, 1047	0.1756	0.8464	0.0245
GINA	Vladimir Nikulin	vn2	1023	0.0226	0.9777	0.0385
HIVA	Chloe Azencott	SVM	992	0.2693	0.7643	0.008
NOVA	Jorge Sueiras	Boost mix	915	0.0659	0.9712	0.3974
SYLVA	Roman Lutz	Doubleboost	893	0.0043	0.9957	0.005