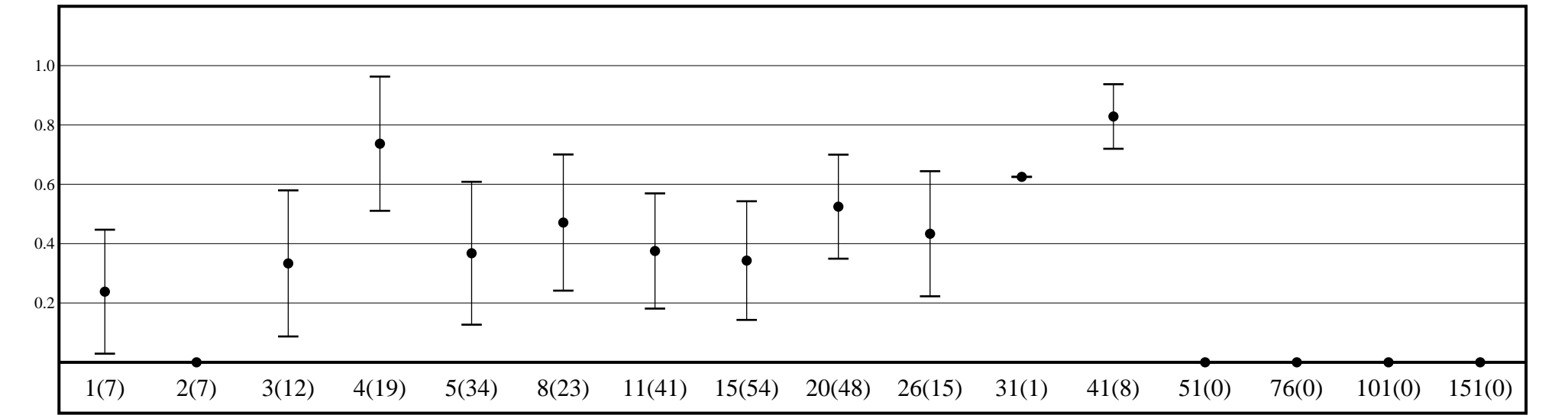


# Test Results Categorized by Training Count

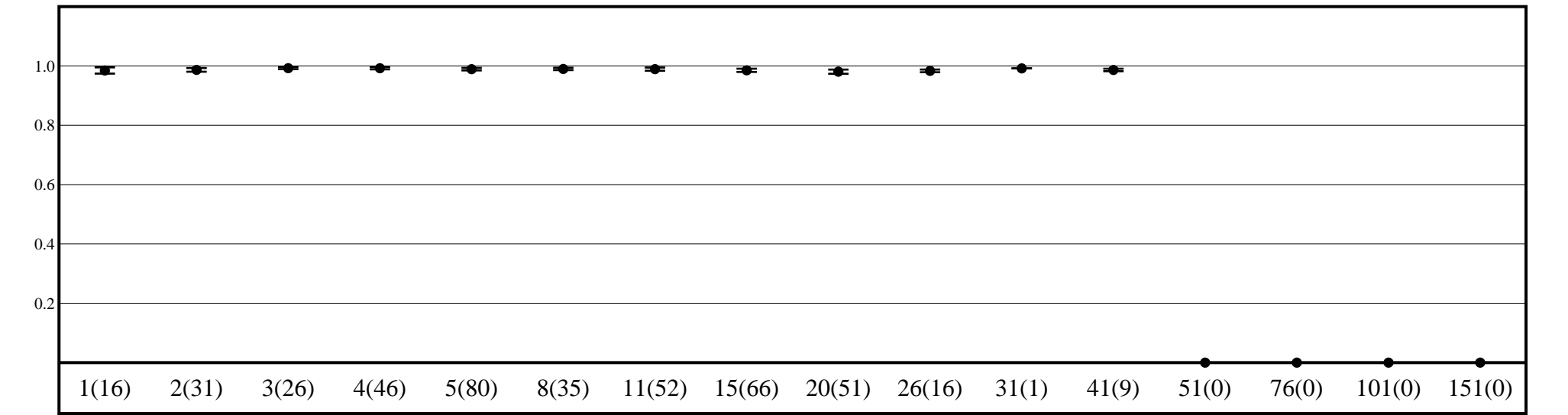
29238380: Set2 100% (UWS=0.45, bias=1.1)

These graphs display the cross validation test results for a classifier arranged by the number of items a classifier was trained on for a given tag. The X axis displays the number of training items, where there is a gap, the label defines the lower bound of a range that extends to the next label. The number in brackets is the number of tags that fall into that 'bin'. The Y Axis is the mean true positive or true negative rate for the classifiers that fall in the 'bin'. Standard Deviation is shown as error bars on the chart.

## True Positive Results



## True Negative Results



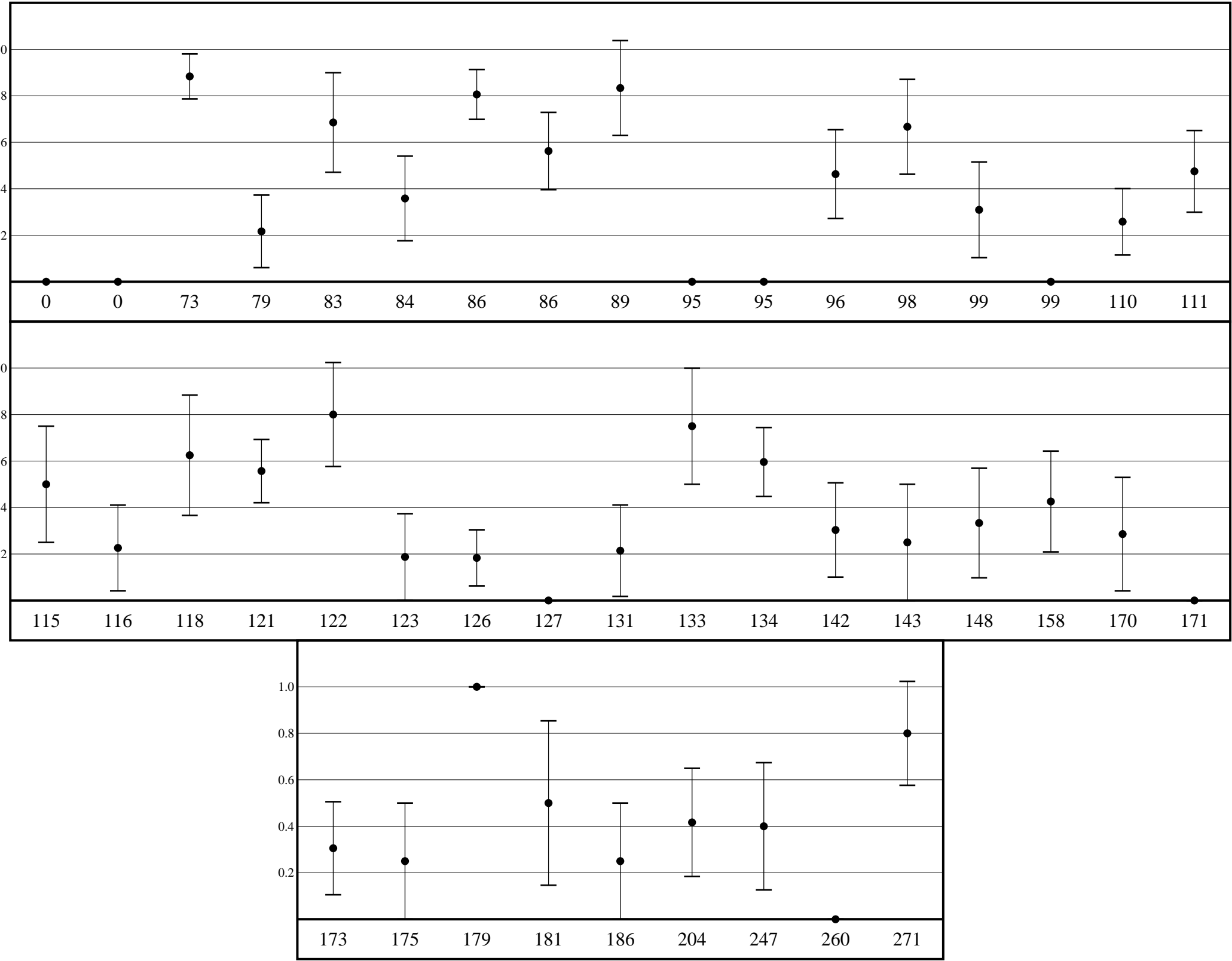
# Tags in bins

Bin	Tags
1..1	ande-spam, ande-publishing, ande-privacy, ande-*missing content, ande-cats,
2..2	ande-language, ande-astronomy, ande-publishing, ande-baseball, ande-*missing content, ande-cats,
3..3	ande-ancient world, ande-language, ande-tags, ande-endangered species, ande-astronomy, ande-baseball, ande-privacy,
4..4	ande-ancient world, ande-recipe, ande-the media, ande-language, ande-tags, ande-endangered species, ande-open source, ande-privacy, ande-solar energy,
5..7	ande-recipe, ande-philosophical, ande-the media, ande-humor, ande-curiosities, ande-tags, ande-reading, ande-evolution, ande-open source, ande-nature photography, ande-antiscience, ande-solar energy,
8..10	ande-philosophical, ande-earth, ande-brain, ande-book review, ande-evolution, ande-antiscience,
11..14	ande-writing, ande-online culture, ande-biology, ande-Vista, ande-health & medicine, ande-earth, ande-book review, ande-music, ande-collaboration, ande-design, ande-*character encoding,
15..19	ande-writing, ande-online culture, ande-biology, ande-education, ande-Vista, ande-health & medicine, ande-green, ande-music, ande-collaboration, ande-food, ande-design,
20..25	ande-social networking, ande-Web 2.0, ande-Google, ande-Internet, ande-green, ande-food, ande-Mac,
26..30	ande-social networking, ande-Web 2.0, ande-Internet,
31..40	ande-climate change,
41..50	ande-climate change,
51..75	
76..100	
101..150	
151..1000	

# Individual Tag Scores

Tag	Instances	Ave Item Size (Distinct Tokens)	Ave Distinct Tokens in Pool	TP	TN
ande-spam	0	0	0	0.00 (0.00)	0.99 (0.01)
ande-*missing content	2	0	0	0.00 (0.00)	1.00 (0.00)
ande-publishing	2	181	181	0.50 (0.71)	0.98 (0.01)
ande-cats	2	95	95	0.00 (0.00)	0.98 (0.01)
ande-astronomy	3	99	197	0.00 (0.00)	0.98 (0.01)
ande-baseball	3	260	520	0.00 (0.00)	0.99 (0.01)
ande-endangered species	4	143	429	0.25 (0.50)	1.00 (0.00)
ande-privacy	4	148	347	0.33 (0.47)	0.99 (0.03)
ande-ancient world	4	133	400	0.75 (0.50)	0.99 (0.01)
ande-language	4	171	478	0.00 (0.00)	1.00 (0.00)
ande-tags	5	186	686	0.25 (0.50)	0.99 (0.01)
ande-open source	5	122	489	0.80 (0.45)	0.98 (0.01)
ande-solar energy	5	271	1084	0.80 (0.45)	0.99 (0.01)
ande-recipe	5	179	715	1.00 (0.00)	0.99 (0.01)
ande-the media	7	175	905	0.25 (0.50)	0.99 (0.01)
ande-curiosities	7	95	531	0.00 (0.00)	0.98 (0.01)
ande-humor	7	127	710	0.00 (0.00)	0.99 (0.01)
ande-nature photography	7	89	518	0.83 (0.41)	0.99 (0.00)
ande-reading	7	247	1367	0.40 (0.55)	0.99 (0.01)
ande-philosophical	8	170	1161	0.29 (0.49)	1.00 (0.00)
ande-antiscience	9	115	893	0.50 (0.50)	0.99 (0.01)
ande-evolution	10	98	815	0.67 (0.41)	0.98 (0.01)
ande-brain	10	118	1031	0.62 (0.52)	0.99 (0.01)
ande-earth	13	173	1879	0.31 (0.40)	0.99 (0.01)
ande-book review	14	204	2362	0.42 (0.47)	0.99 (0.01)
ande-*character encoding	14	131	1574	0.21 (0.39)	1.00 (0.00)
ande-writing	15	123	1635	0.19 (0.37)	0.99 (0.01)
ande-health & medicine	16	86	1214	0.56 (0.33)	0.98 (0.01)
ande-collaboration	16	116	1599	0.23 (0.37)	0.99 (0.01)
ande-design	17	158	2395	0.43 (0.43)	0.98 (0.01)
ande-music	17	110	1640	0.26 (0.29)	0.99 (0.01)
ande-Vista	18	83	1338	0.69 (0.43)	0.97 (0.01)
ande-education	18	126	2051	0.18 (0.24)	0.99 (0.00)
ande-online culture	18	142	2304	0.30 (0.41)	0.99 (0.01)
ande-biology	19	99	1664	0.31 (0.41)	0.99 (0.00)
ande-food	23	96	1971	0.46 (0.38)	0.99 (0.01)
ande-green	24	134	2822	0.60 (0.30)	0.99 (0.01)
ande-Google	25	84	1906	0.36 (0.36)	0.98 (0.01)
ande-Mac	26	73	1708	0.88 (0.19)	0.96 (0.01)
ande-Web 2.0	28	121	3021	0.56 (0.27)	0.98 (0.01)
ande-Internet	28	79	2007	0.22 (0.31)	0.99 (0.01)
ande-social networking	29	111	2912	0.47 (0.35)	0.98 (0.01)
ande-climate change	48	86	3691	0.81 (0.21)	0.99 (0.01)
Average				0.397073255228017	0.987513008575268

True Positive Results by Average Distinct Tokens per Item



True Positive Results by Average Distinct Token Count in Pool

