

Appendix

This is the Appendix to the paper *FactCheck: Validating RDF Triples using Textual Evidence*.

Tables 6 and 7 show the results of FACTCHECK on all FactBench sub sets with ClueWeb and Wikipedia as reference corpora, respectively. Tables 8 and 9 show the results of DEFACTO on all FactBench sub sets with ClueWeb and Wikipedia as reference corpora, respectively. Table 10 shows the results of FACTCHECK when using various feature sub set evaluation techniques.

Table 6: Classification results of FACTCHECK on ClueWeb corpus

	Domain					Range				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.917	0.913	0.913	0.919	0.279	0.922	0.917	0.917	0.922	0.272
Simple Logistics	0.910	0.900	0.900	0.951	0.272	0.915	0.910	0.910	0.956	0.264
Naive Bayes	0.827	0.790	0.790	0.901	0.445	0.833	0.800	0.790	0.906	0.440
SMO	0.905	0.900	0.900	0.900	0.315	0.910	0.900	0.900	0.905	0.309
	Domain-Range					Random				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.923	0.919	0.919	0.924	0.275	0.931	0.920	0.920	0.935	0.257
Simple Logistics	0.911	0.907	0.907	0.951	0.271	0.930	0.920	0.920	0.960	0.247
Naive Bayes	0.822	0.790	0.785	0.900	0.451	0.852	0.830	0.830	0.921	0.399
SMO	0.905	0.900	0.900	0.901	0.315	0.921	0.915	0.915	0.915	0.291
	Mix					Property				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.891	0.885	0.885	0.925	0.299	0.819	0.819	0.819	0.865	0.374
Simple Logistics	0.893	0.883	0.883	0.955	0.288	0.806	0.805	0.805	0.860	0.381
Naive Bayes	0.808	0.760	0.760	0.869	0.475	0.677	0.660	0.660	0.747	0.551
SMO	0.885	0.871	0.871	0.877	0.358	0.812	0.812	0.812	0.812	0.434

Table 7: Classification results of FACTCHECK on Wikipedia corpus

	Domain					Range				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.856	0.852	0.851	0.856	0.352	0.855	0.850	0.851	0.856	0.354
Simple Logistics	0.855	0.853	0.853	0.920	0.328	0.853	0.851	0.851	0.921	0.328
Naive Bayes	0.803	0.723	0.703	0.861	0.519	0.807	0.725	0.705	0.863	0.517
SMO	0.840	0.840	0.840	0.840	0.400	0.835	0.835	0.835	0.835	0.406
	Domain-Range					Random				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.860	0.855	0.854	0.862	0.350	0.859	0.854	0.853	0.858	0.353
Simple Logistics	0.846	0.845	0.845	0.919	0.332	0.846	0.834	0.833	0.917	0.328
Naive Bayes	0.800	0.721	0.701	0.861	0.520	0.814	0.762	0.751	0.904	0.483
SMO	0.832	0.832	0.832	0.832	0.400	0.840	0.822	0.820	0.823	0.421
	Mix					Property				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.840	0.837	0.837	0.825	0.391	0.778	0.773	0.772	0.816	0.424
Simple Logistics	0.836	0.835	0.835	0.883	0.368	0.663	0.663	0.663	0.756	0.445
Naive Bayes	0.768	0.684	0.659	0.894	0.555	0.645	0.591	0.550	0.685	0.597
SMO	0.826	0.825	0.825	0.825	0.4187	0.678	0.675	0.674	0.729	0.456

Table 8: Classification results of DEFACTO on ClueWeb corpus

	Domain					Range				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.861	0.861	0.861	0.863	0.350	0.864	0.864	0.864	0.866	0.347
Simple Logistics	0.855	0.840	0.840	0.901	0.348	0.858	0.858	0.858	0.903	0.344
Naive Bayes	0.806	0.780	0.780	0.897	0.452	0.808	0.780	0.780	0.899	0.448
SMO	0.855	0.854	0.854	0.854	0.381	0.858	0.857	0.857	0.857	0.378
	Domain-Range					Random				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.867	0.867	0.867	0.869	0.343	0.872	0.872	0.872	0.879	0.339
Simple Logistics	0.860	0.860	0.860	0.904	0.340	0.866	0.866	0.866	0.913	0.334
Naive Bayes	0.811	0.790	0.780	0.903	0.440	0.820	0.805	0.803	0.909	0.430
SMO	0.860	0.860	0.860	0.904	0.340	0.862	0.862	0.862	0.862	0.371
	Mix					Property				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.832	0.830	0.830	0.829	0.384	0.713	0.703	0.699	0.758	0.452
Simple Logistics	0.819	0.818	0.818	0.871	0.378	0.709	0.700	0.700	0.781	0.441
Naive Bayes	0.762	0.740	0.740	0.863	0.480	0.661	0.654	0.649	0.724	0.519
SMO	0.821	0.820	0.820	0.816	0.424	0.720	0.710	0.710	0.719	0.530

Table 9: Classification results of DEFACTO on Wikipedia corpus

	Domain					Range				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.840	0.831	0.830	0.836	0.367	0.845	0.834	0.833	0.840	0.364
Simple Logistics	0.836	0.820	0.818	0.842	0.378	0.842	0.824	0.822	0.845	0.376
Naive Bayes	0.767	0.699	0.678	0.829	0.536	0.780	0.705	0.683	0.837	0.530
SMO	0.829	0.809	0.807	0.809	0.436	0.834	0.813	0.810	0.812	0.432
	Domain-Range					Random				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.839	0.830	0.829	0.837	0.366	0.840	0.838	0.838	0.844	0.364
Simple Logistics	0.838	0.822	0.819	0.847	0.376	0.807	0.807	0.807	0.887	0.371
Naive Bayes	0.771	0.701	0.679	0.828	0.533	0.797	0.743	0.732	0.881	0.499
SMO	0.832	0.811	0.808	0.811	0.434	0.814	0.814	0.814	0.814	0.431
	Mix					Property				
	P	R	F1	AUC	RMSE	P	R	F1	AUC	RMSE
J48	0.809	0.802	0.803	0.827	0.392	0.753	0.752	0.752	0.805	0.422
Simple Logistics	0.795	0.784	0.784	0.847	0.394	0.737	0.734	0.733	0.811	0.422
Naive Bayes	0.726	0.600	0.554	0.820	0.620	0.654	0.628	0.611	0.692	0.539
SMO	0.787	0.777	0.777	0.782	0.471	0.657	0.657	0.657	0.657	0.585

Table 10: Classification sub-set results for FACTCHECK

	J48		Simple Logistics	
	F-Measure	AUC	F-Measure	AUC
WrapperSubsetEval	0.898	0.94	0.876	0.928
ClassifierSubsetEval	0.898	0.923	0.872	0.936
FilteredSubsetEval	0.898	0.923	0.872	0.936
CfsSubsetEval	0.884	0.910	0.869	0.93
ConsistencySubsetEval	0.886	0.914	0.892	0.955