Using COBRA for Classification tasks on Imbalanced Data

MA 691
Advanced Statistical Algorithms

by

Vishisht Priyadarshi (180123053)
Aadi Gupta (180123059)
Tejus Singla (180123061)
Shashank Goyal (180123042)

Indian Institute of Technology Guwahati, Assam

Dataset - Red wine quality

Undersampling Method	Accuracy	Precision	Recall	F1 - Score
None (18 vs 1500)	0.9931195244055069	0.9965464043622705	0.6944444444444444	0.8185095911415762
Near Miss v1 (18 vs 80)	0.9975	0.9987421383647799	0.8888888888888888	0.9406189841551903
Near Miss v2 (18 vs 80)	0.9949976533166458	0.9974838727052666	0.7777777777777777	0.8740354298430038
Near Miss v3 (18 vs 80)	0.9975	0.9987421383647799	0.8888888888888888	0.9406189841551903
Condensed KNN (18 vs 64)	0.9843726533166458	0.8176954874901351	0.882251720747296	0.8487478286263477
KNN Und (18 vs 1497)	0.9931179599499375	0.7465534629510091	0.6944444444444444	0.7195567768230132
Edited KNN (18 vs 38)	0.9943726533166458	0.9971721667476239	0.75	0.8561023799422142

Dataset - Car Evaluation

Undersampling Method	Accuracy	Precision	Recall	F1 - Score
None (65 vs 1728)	0.9762731481481481	0.967417149958575	0.6924074909747293	0.8071296991374843
Near Miss v1 (65 vs 164)	0.9866898148148149	0.9825983590607446	0.8304283243080626	0.900127412649535
Near Miss v2 (65 vs 164)	0.9895833333333333	0.9946463960138158	0.8617424242424243	0.9234369300356218
Near Miss v3 (65 vs 164)	0.9884259259259259	0.9940634726513735	0.8468276515151515	0.9145575476480056
Condensed KNN (65 vs 189)	0.9513888888888888	0.7546303142669006	0.7973098934008229	0.7753832427982211
KNN Und (65 vs 1324)	0.9849537037037037	0.8910826121268989	0.9100251574162901	0.9004542738465549
Edited KNN (65 vs 160)	0.9890046296296295	0.9842261904761905	0.8614415818838201	0.9187497112453408

Dataset - Ecoli

Undersampling Method	Accuracy	Precision	Recall	F1 - Score
None (35 vs 336)	0.8720238095238095	0.5409226190476191	0.5911569925983639	0.5649252674724745
Near Miss v1 (35 vs 157)	0.9107142857142858	0.6227197608114245	0.6648811842617842	0.6431102021410434
Near Miss v2 (35 vs 157)	0.9047619047619048	0.7557714809254681	0.6966026057222006	0.7249817905501151
Near Miss v3 (35 vs 157)	0.9077380952380953	0.7495791245791246	0.660741462147773	0.7023623017789314
Condensed KNN (35 vs 80)	0.8184523809523809	0.6603370649190206	0.7583023849716488	0.7059371868614849
KNN Und (35 vs 197)	0.8511904761904762	0.7007020757020757	0.8908862485391507	0.7844312930146947
Edited KNN (35 vs 52)	0.5357142857142857	0.5385816738458762	0.5519633813790417	0.5451904264211119

Dataset - Abalone

Undersampling Method	Accuracy	Precision	Recall	F1 - Score
None (62 vs 4177)	0.9851568102577082	0.4925784051288541	0.5	0.4962614566099781
Near Miss v1 (62 vs 278)	0.9712745699513416	0.5305069635376267	0.5882707921878428	0.5578976702107881
Near Miss v2 (62 vs 278)	0.6114961557437333	0.5027193282069707	0.5089304583470284	0.5058058263340073
Near Miss v3 (62 vs 278)	0.6358185047016942	0.4961538469149962	0.4180150275443231	0.453744969400888
Condensed KNN (62 vs 155)	0.9815662318035174	0.4925516303375210	0.4981776653990994	0.4953486736407133
KNN Und (62 vs 3808)	0.9837207164329117	0.5239342776553418	0.5072141759929778	0.5154386683608703
Edited KNN (62 vs 79)	0.4913923782484057	0.50592290607116	0.5115156756993844	0.5087039194060129

Dataset - Nursery

Undersampling Method	Accuracy	Precision	Recall	F1 - Score
None (328 vs 12960)	0.9746913580246913	0.4873456790123456	0.5	0.4935917474210691
Near Miss v1 (328 vs 1476)	0.9746913580246913	0.6124189790985137	0.5029696164599391	0.5523243473289116
Near Miss v2 (328 vs 1476)	0.9392746913580247	0.6523999928787515	0.7995808168183012	0.7185308727421603
Near Miss v3 (328 vs 1476)	0.9746913580246913	0.4873456790123456	0.5	0.4935917474210691
Condensed KNN (328 vs 1171)	0.9756172839506172	0.6879452853266063	0.5242319158467075	0.5950332584631954
KNN Und (328 vs 10071)	0.9746913580246913	0.4873456790123456	0.5	0.4935917474210691
Edited KNN (328 vs 418)	0.4909722222222222	0.5231155011330089	0.731453412935016	0.6099858116557852

Thank You

Thanks for your attention