Your two personalized problems can be found in the table in "Project2\_Fall\_2022.pdf" (a smaller version of that table is below, it may be hard to read)

To help you check your code, I am going to give you the answers to the last three datasets.

It is important to note (as I will explain in more detail in class), you may get slightly different results to the gold standard below.

However, you should:

- Get 2 out of 3 out correct features. For example, for Sue's small dataset, the correct answer is {1,5,4}, but if Sue got {1,4}, that would be correct, even though she is missing feature "5"
- For Sue's small dataset, the correct accuracy is 0.922. However, if Sue got within plus or minus 3% of that, that would be considered correct.

For Sue (who needs to work on small datasets 96 and large dataset 21)

On small dataset 123 the error rate can be 0.922 when using only features 1 5 4

On large dataset 123 the error rate can be 0.954 when using only features 7 11 14 For Joe

On small dataset 124 the error rate can be 0.940 when using only features 5 2 3

On large dataset 124 the error rate can be 0.941 when using only features 39 5 13

On small dataset 125 the error rate can be 0.912 when using only features 1 4 2

On large dataset 125 the error rate can be 0.946 when using only features 18 25 34

SID	LAST NAME	FIRST NAME	Small Large	66	7814 M	Ar	32	90
	5249 Al	CP	81 118	67	7689 M	In	99	32
	0030 Ar	O	105 81	68	4812 M	Ra	64	99
	7833 Ar	Lu	53 105	69	9022 Na	Ja	109	64
	7884 Ar	Lit	79 53	70	7874 Na	Re	5	109
	0805 Ba	Hi Cr	94 79	71	6909 No	Al	16	5
	0322 Ba	Cr	104 94	72	6910 No	Za	10	16
	5321 BF	E .	48 104	73	2778 Ot	Tr	106	10
-	6336 Bc	Al Ju	38 48	74	4043 Pa	21	33	106
	4471 Bd 3453 Bd	10	11 38 39 11	75	7380 Pa	Br	61	33
	1115 Bu	Ar Fr	59 11 67 59	76	1854 Pa	Ju	78	61
	1761 Ca	Jo	24 67	77	3965 PH	H	46	78
	1945 CH	19	74 24	78	8256 PH	Ar	4	46
	2333 Ch	Ja N	103 74	79	2843 PH	Et	108	4
	0396 CH	Be	91 103	80	7794 Pc	Ez	8	108
	5214 CH	Cr.	23 91	81	1940 Pr	Ke	122	8
	6340 Ch	A	28 23			Cr		
	3516 Cd	Ar Al	97 28	82	5850 Ra		113	122
	8465 Cc	Di	36 97	83	9641 Ra	Ad	65	113
	8616 Da	Za	70 36	84	7355 Ra	Al	56	65
	5551 De	Ja Er	116 70	85	7957 Ra	Da	85	56
	8565 De	Er	15 116	86	2100 Rd	N:	100	85
	2286 De	Ca	93 15	87	3417 Ru	Ze	114	100
	4795 De	Ct M	57 93	88	8714 Sa	Br	27	114
	3921 De	Hi M	77 57	89	5098 Sa	Sk	110	27
	1078 Fu	M	73 77	90	3625 Sa	Is	121	110
	6117 G	Ar	49 73	91	7096 Sa	N	84	121
	5609 Ga	Jo	112 49	92	7285 SH	Br	75	84
	6679 GI	Ju	2 112	93	9033 SH	Al	45	75
	2487 G	Di	82 2	94	8895 SH	Ra	102	45
	9874 Hz	K)	87 82	95	8900 SH	Di	66	102
	4078 Hi	D	76 87	96	6109 SH	H	- 50	66
	6392 He	Ju	72 76	97	0103 SF	ca	86	50
	1318 He	CI	1 72	98	8533 Si	M	68	86
	1603 Hr	Al	98 1	99	3701 Sr	Se	123	68
	0164 H	Jin	117 98	100	3492 Ta	Ja Ja	18	123
	2398 Hu	At At	119 117			Al		
	7512 Hi 6343 Id	Ke	69 119	101	5893 TH	Je	115	18
	5931 In	Ju	89 69 83 89	102	5936 Ti		3	115
	3588 Is	St	62 83	103	9395 Tc	Ri	19	3
	2815 Ja	M	40 62	104	7455 To	Va	7	19
	2375 Jo	M	54 40	105	9565 Tr	Ju	63	7
	6673 Ke	Er	25 54	106	5356 Tu	Ju	34	63
	6702 Ka	Re	71 25	107	2585 VI	Ri	9	34
	8343 Ki	M	52 71	108	7364 Vo	M	37	9
	3293 Kg	N	43 52	109	7045 Vu	Ri	80	37
	4132 Ku	Jo	29 43	110	6532 Vu	Cr.	47	80
	3377 Le	н	30 29	111	6246 W	At	58	47
	9014 Le	Hi Gi	20 30	112	7611 W	Zr	13	58
	2209 Lis	Ke	35 20	113	6463 W	Or .	41	13
	2222 Lis	Q	120 35	114	8683 W	De	59	41
	5458 Lis	Br	22 120	115	5068 W	Re	124	59
	6646 Lit	A	95 22	116	7627 W	Te	107	124
	7832 Lin	Di	17 95	117	9826 Xi	Ed	31	107
	1776 Lo		101 17	118	3679 Xu	Je Eq	125	
	3384 M	Ja Vi	14 101					31
	2182 M	Ar	111 14	119	4632 Xu	Ze	92	125
	5676 M	Jo	55 111	120	7553 Yd	M	60	92
	9627 M	Di	51 55	121	2938 Ze	Ra	26	60
	6756 M	Tr	44 51	122	7767 Zu	Sue	21	26
	1538 M	U:	42 44		123 Sample data		96	21
	6780 M	Ar	12 42	124 Sample dat		Joe	6	96
	7065 M	Ar	90 12	125 Sample dat	a	Van	88	6