Weekly Research Project Status Report

Project Title : Stellar spectral classification using Active learning approach
Report Date : 8th March, 2025

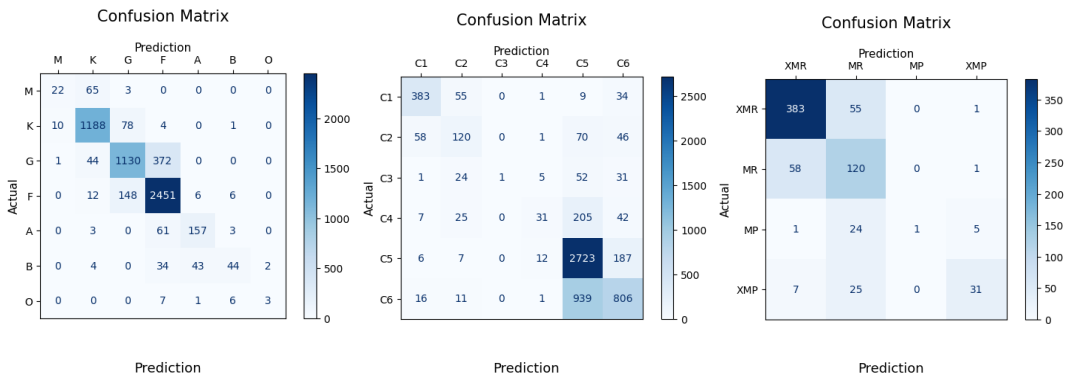
Summary of Progress

- We have completed the preprocessing of the data

	manga_id	teff	teff_err	logg	logg_err	feh	feh_err	flux	s2n	teff_class	teff_label	logg_label	feh_class	feh_label
0	13-0	6197.079590	75.208549	4.284935	0.141657	-1.81075	0.229987	[277.32285, 289.5346, 299.2491, 285.91418, 284...	126.844055	F	3	4	MP	1
1	13-0	6197.079590	75.208549	4.284935	0.141657	-1.81075	0.229987	[295.3848, 315.6807, 310.15118, 303.53024, 293...	93.818542	F	3	4	MP	1
2	13-0	6197.079590	75.208549	4.284935	0.141657	-1.81075	0.229987	[266.91608, 271.9149, 281.03305, 280.8652, 264...	125.689880	F	3	4	MP	1
3	13-1	6183.310547	88.320694	4.335517	0.221244	-1.39514	0.235241	[172.67696, 171.515, 170.73587, 167.45573, 168...	96.085899	F	3	4	MP	1
4	13-1	6183.310547	88.320694	4.335517	0.221244	-1.39514	0.235241	[155.72183, 175.82281, 164.6047, 156.97221, 16...	71.113739	F	3	4	MP	1

Figure 1 : The head of the final dataframe at the end of preprocessing

- Completed Feature selection process (we reduced the features from 4563 to 170)
- Applied PCA on 170 features. We found 9 PCA components covering 99.95% variance.
- We performed simple k-NN without weights and did classification in terms of T_eff, log(g) and [Fe/H] (k = 230, training size = 53176, test size = 5909)
- We performed simple k-NN with weights and did classification in terms of T_eff, log(g) and [Fe/H] (k = 230, training size = 53176, test size = 5909)



- We found that k-NN with weights performs better than k-NN without weights. (k = 230, training size = 53176, test size = 5909)

Planned Activities for Next Week

- Try to implement simplest Active Learning Algorithm
- Learn dimensionality reduction methods other than PCA