Weekly Research Project Status Report

Project Title : Stellar spectral classification using Active learning approach
Report Date : 1st 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

- We will be classifying the stars in three ways :

  1. Using their Teff (Surface Temperature) => classes : O,B,A,F,G,K,M
  2. Using log(g) (Surface gravity) => classes : 1,2,3,4,5,6
  3. Using Fe/H (metallicity) => classes : XMR, MR, MP, XMP

- We have created the labels for each of the above mentioned classification (i.e. teff_label,log_label and feh_label, respectively)
- Following is the plot representative of our data after the end of the preprocessing. Each point in the plot represents an individual star. Here, We have taken stars who have S/N > 50.

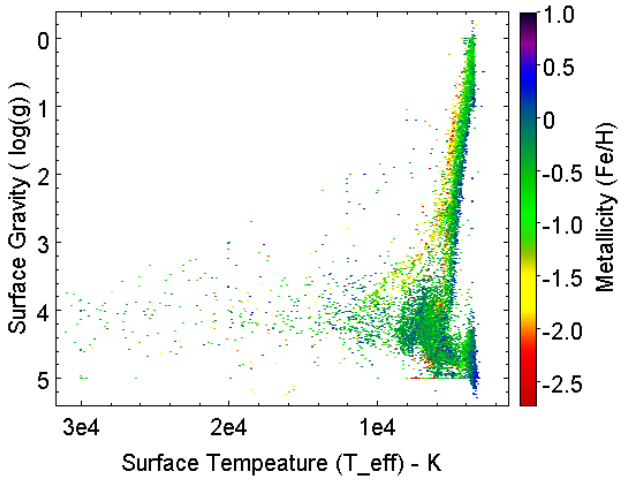

Figure -2 : The representative plot of our refined dataset as mentioned above

- Learnt Principal Component Analysis (PCA) and implemented on a simple example

- A little progress in feature selection

  We are reading about how to select good features in our case i.e. spectral features

Planned Activities for Next Week

- Completing the feature selection
- Active learning implementation on an example dataset