Assignment 1
Milestone 3
Group 1

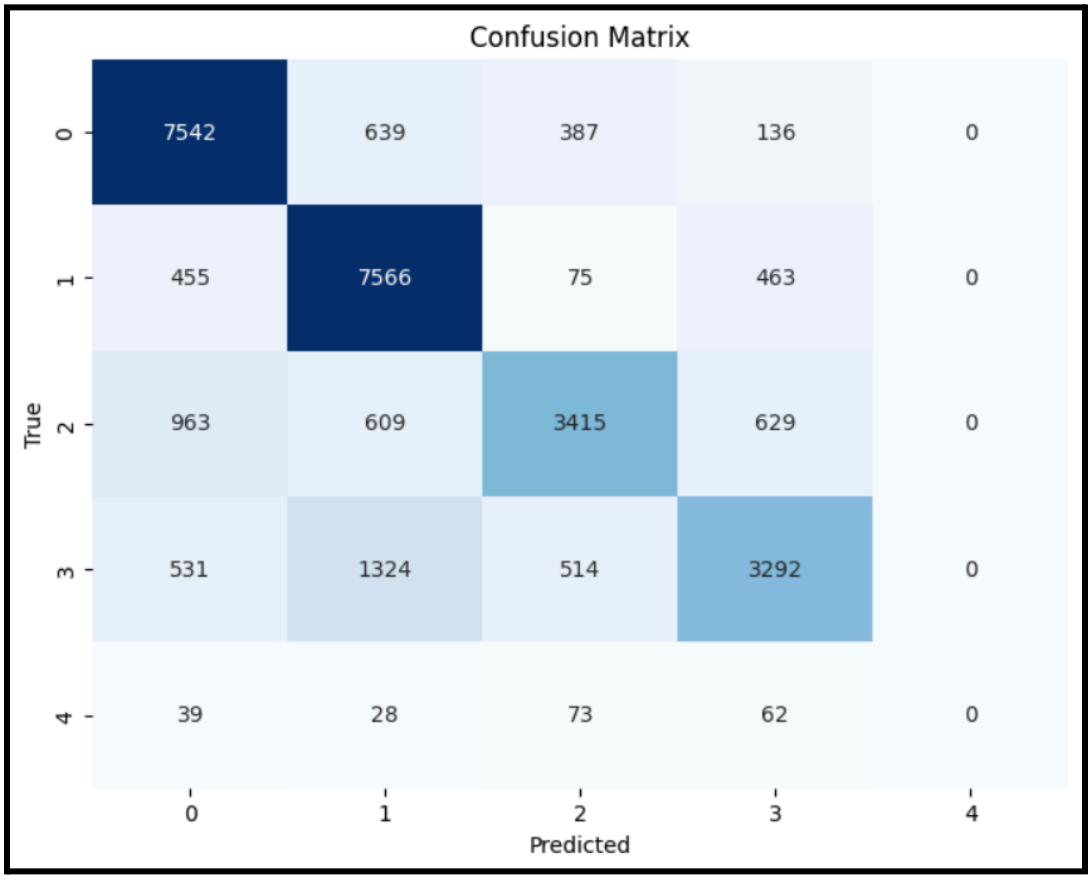
Aaditya Gupta **Harjeet Singh Yadav**

Selective Ensemble on data with noisy features and labels i.e. predicting target_10_val:

The Ensemble:

- Contains 12 simple NN models
- Each model has a different combination of activation functions, optimizers and batch size
- Each model gives out a probability of classifying a test point to a class during testing.
- If this probability is more than a threshold value(taken as 0.56), then it is considered for comparison, else not.
- All 12 probabilities, for each test point are compared, and the probability which is maximum is considered, and that label is considered as the final prediction.

Confusion matrix:



Where 0:0, 1:0.25, 2:0.5, 3:0.75, 4:1.0

Classification Report:

	precision	recall	f1-score	support
0	0.79	0.87	0.83	8704
1	0.74	0.88	0.81	8559
2	0.77	0.61	0.68	5616
3	0.72	0.58	0.64	5661
4	0.00	0.00	0.00	202
accuracy			0.76	28742
macro avg	0.60	0.59	0.59	28742
weighted avg	0.75	0.76	0.75	28742

The reason that class 4 is not predicted at all might be that the number of sample with label “1” are only 2k, as compared to other class members being around 60k.

More experiments in Ensemble on predicting ‘era’

- **RandomForestClassifier**
- Zero noise:- Training accuracy = 0.99, test accuracy = 0.96
- Low Noise:- Training accuracy = 0.99, test accuracy = 0.83
- High Noise:- Training accuracy = 1.0 , test accuracy = 0.6

- GradientBoostingClassifier on high-noise
    - Training accuracy = 0.53, test accuracy = 0.52

- AdaboostClssifier on high-noise
    - Training accuracy = 0.34, test accuracy = 0.34

Cascade/ Conservative learning

The algorithm sequentially trains decision tree models on subsets of data where previous models are confident, determined by the Gini index of predicted class distributions. Training continues until a minimum desired accuracy threshold on a test set is met. During testing,

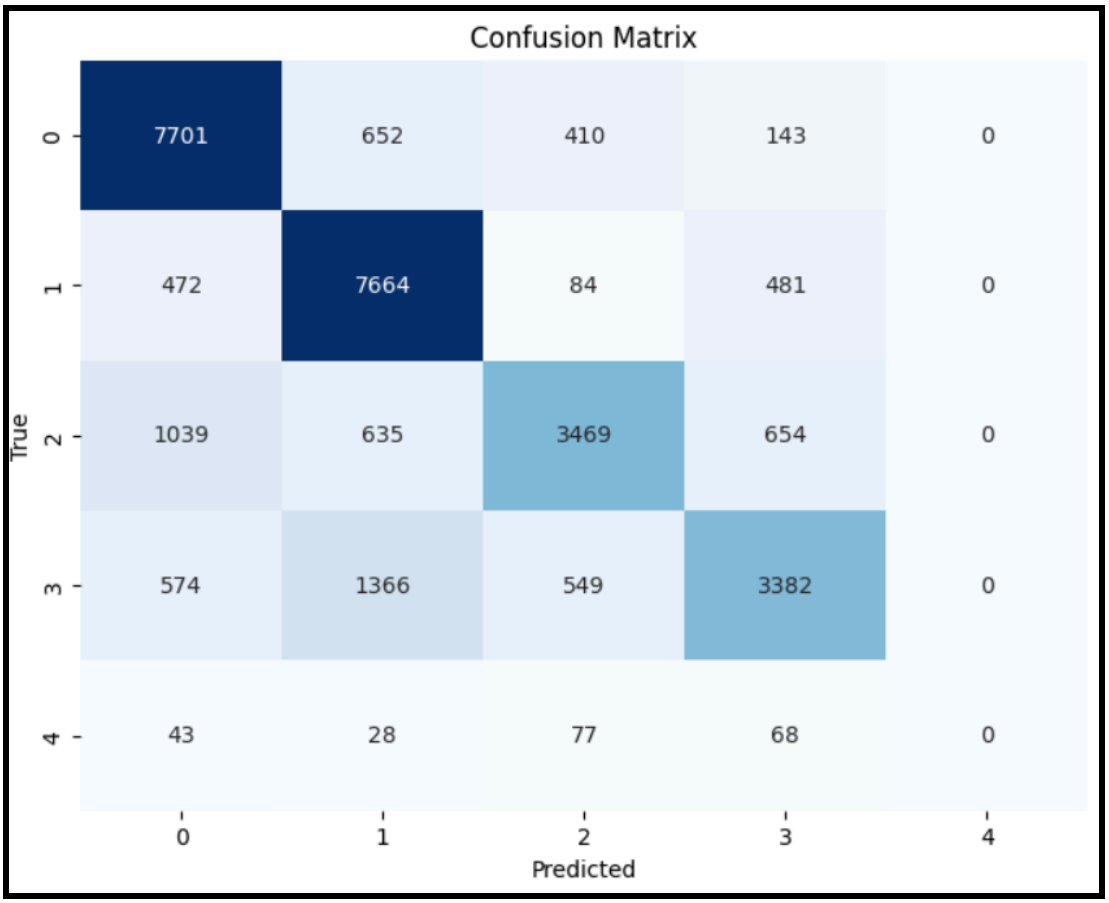
predictions from all models are combined, and the mode prediction across all models is used as the final prediction for each data point. Finally, the overall accuracy of the combined predictions is calculated on the test set.

On low noise:- accuracy:- 0.78
On high noise:- accuracy:- 0.515

Noise mitigation experimentation:

- Applied model cascading to this selective ensemble.
- Trained fresh models on the dataset, and used these models to make predictions on the train set itself.
- The data points for which the entire ensemble couldn't give a prediction with a confidence level > 0.8 were dropped, and sent to the next ensemble.
- These pruned data points were then used as train and test data points for the next fresh ensemble.

Results:



	precision	recall	f1-score	support
0	0.78	0.86	0.82	8906
1	0.74	0.88	0.80	8701
2	0.76	0.60	0.67	5797
3	0.72	0.58	0.64	5871
4	0.00	0.00	0.00	216
accuracy			0.75	29491
macro avg	0.60	0.58	0.59	29491
weighted avg	0.75	0.75	0.74	29491

** This observation is when cascading is done using pruning on **test dataset**, rather than on the **train set.** Model cascading on the training set is still under process, please consider this in the meanwhile**