Comparing Results

In this report, Machine Learning results are compared when we include redshift dependant features amoung the features we obtained from busy function fitting and excluding the redshift dependant features. As usual we have Random Forest, Decision tree, KNN, Logistic Regression and SVM as machine Learning Models. Normal Neural network has failed in both the cases as it requires a large amount of data to train, but CNN showed it can be trained, as it has been overfitting using previous data. We can train CNN by cropping the spectra and feeding it. But how the spectra should be cropped has been an issue as the widths are uneven between Associated and Intervening spectra. During the meetings with Dr Banerjee, we have decided to crop just the spectra i.e only the part where spectra is seen. This work is kind of troublesome, but I can get it done.

Here, we compare ROC AUC, Accuracy and Precision. Just to recap, ROC AUC is the area under the reciever operating characteristic curve. This ROC AUC can be interpreted as the proability that the model predicts a random positive sample more highly than a random negative sample. It should be near 1 to make sure the model is predicted correctly. Average Accuracy and Average precision talks about the accuracy and precision of the model.

1.1 Comapring ROC AUC, Accuracy and Precision

1.1.1 When Redshift dependant features are added

ML Model	ROC AUC —Average Accuracy	Average Precision
Random Forest	0.942	88.9%	92.6%
KNN	0.882	82.1%	85.1%
Decision Tree	0.754	75.5%	67.2%
Logistic Regression	0.854	76.9%	82.7%
SVM	0.842	76.6%	82.4%

Table 1.1: Results for Machine Learning models incluing redshift dependant features.

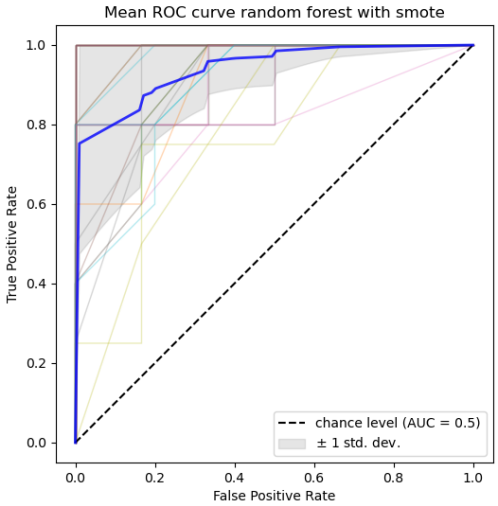
1.1.2 When Redshift dependant features are not added

ML Model	ROC AUC —Average Accuracy	Average Precision
Random Forest	0.864	79.9%	82.5%
KNN	0.775	71.4%	76.8%
Decision Tree	0.812	81.6%	74.9%
Logistic Regression	0.510	46.5%	62.2%
SVM	0.557	48.9%	63.2%

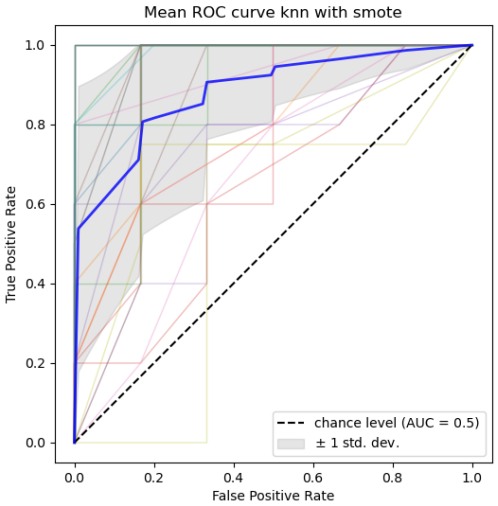
Table 1.2: Results for Machine Learning models not incluing redshift dependant features.

1.2 Comparing the ROC Curves

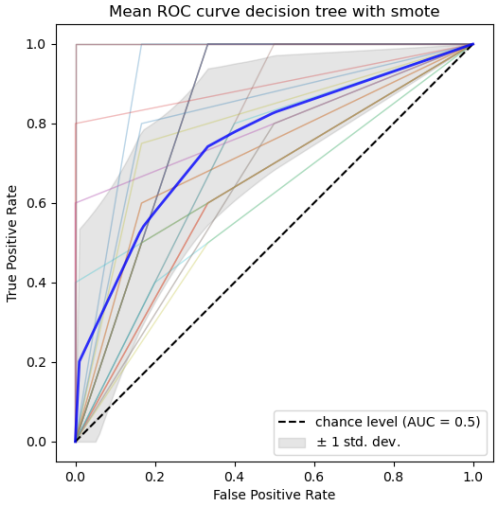
1.2.1 When Redshift dependant features are added



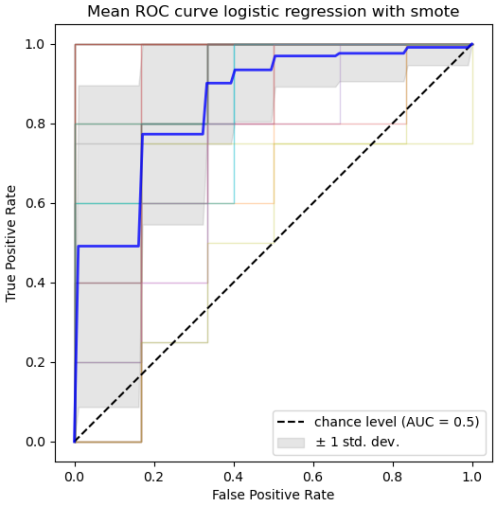
(a) Random Forest



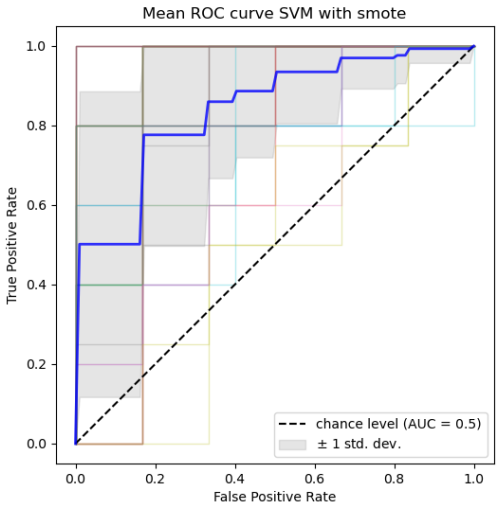
(b) KNN



(c) Decision Tree



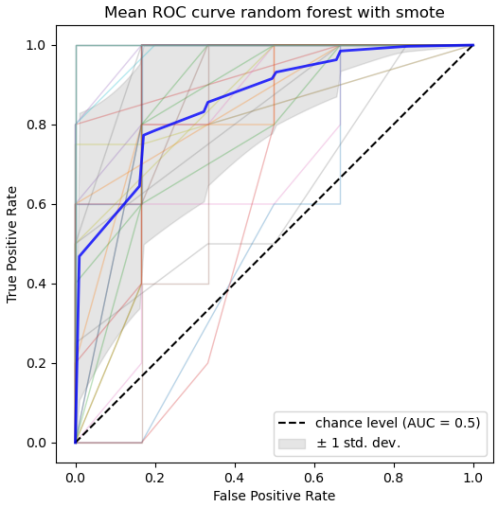
(d) Logistic Regression



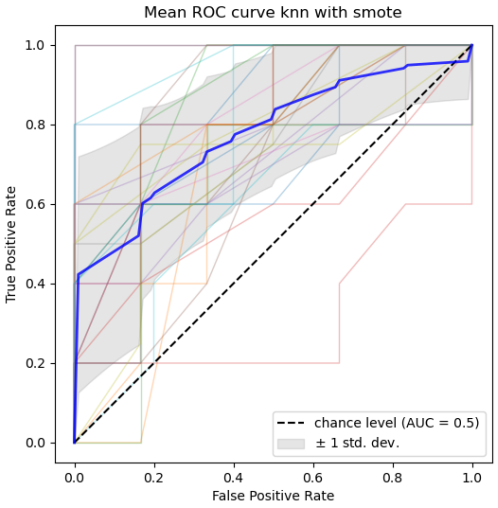
(e) SVM

Figure 1.1: Mean ROC curve plots for all models when redhsift dependant features are added

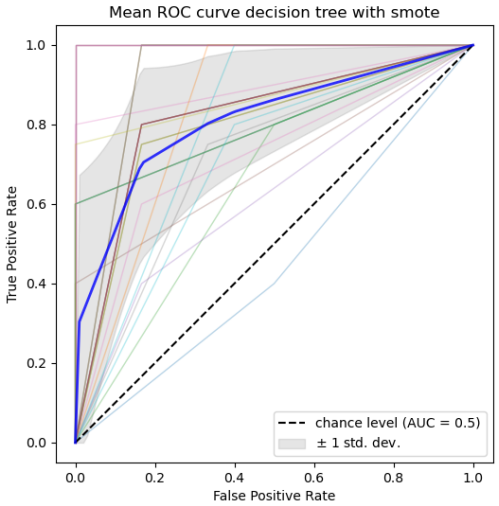
1.2.2 When Redshift dependant features are not added



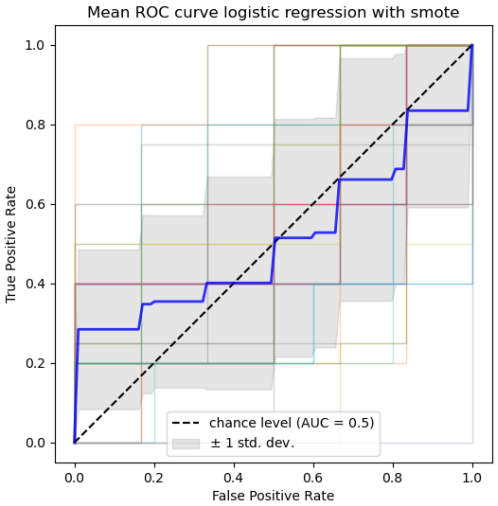
(a) Random Forest



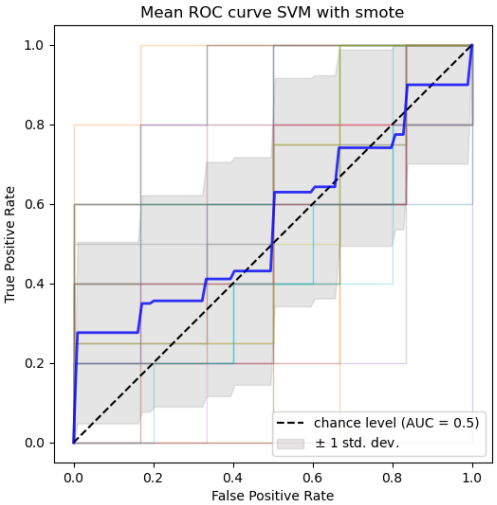
(b) KNN



(c) Decision Tree



(d) Logistic Regression



(e) SVM

Figure 1.2: Mean ROC curve plots for all models when redhsift dependant features are not added