



Predicting gas-rich galaxies using machine learning

By: Mariah Ricci and Brendan Miller



What are we doing?

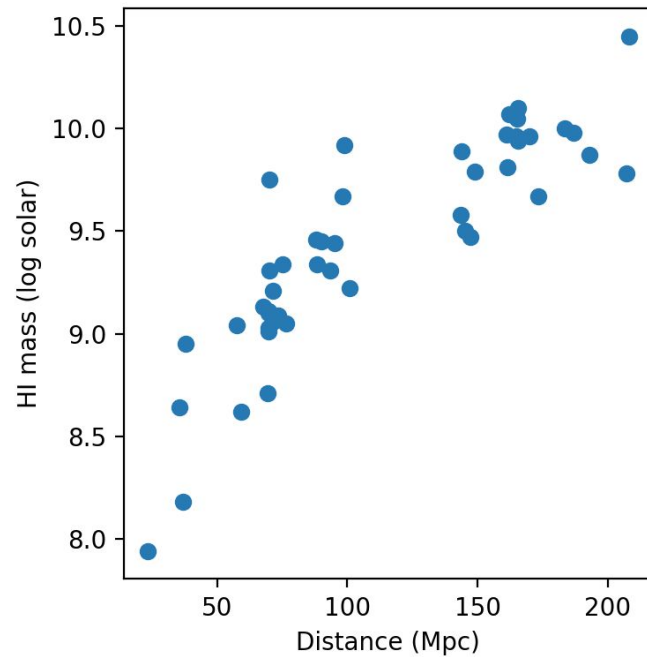
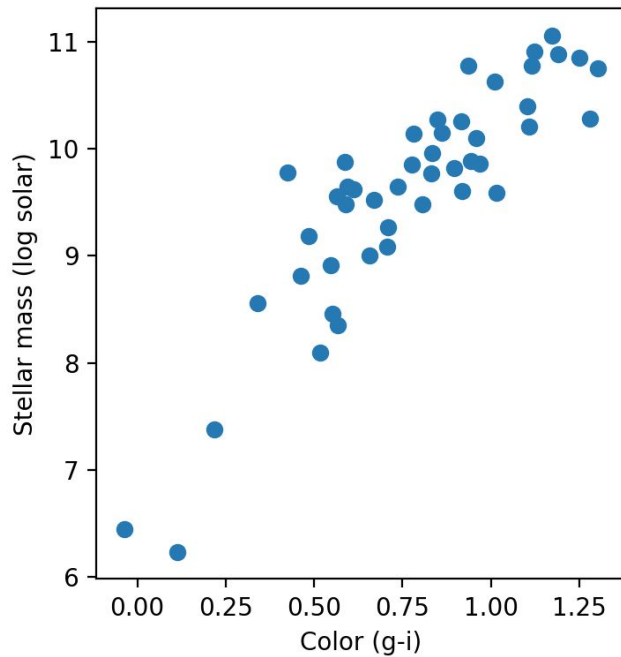
Examining how the HI gas fraction or detection likelihood is related to SDSS (and GALEX, and 2MASS) properties with the end result potentially being useful down the road in target selection.



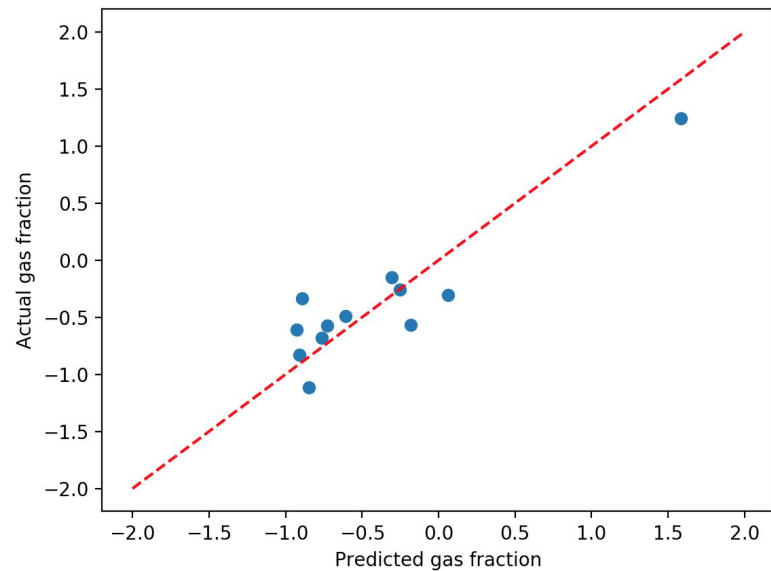
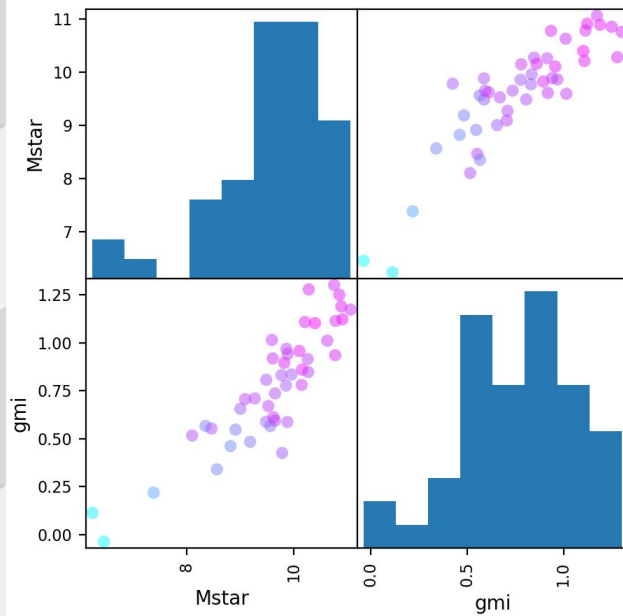
How is it implemented?

- Data from SDSS (and GALEX, and 2MASS)
- HI gas fraction (ALFALFA) or detection (APPSS)
- Use code that takes in a list of properties and returns either a predicted gas fraction (regression) or a yes/no predicted detection (classification)
- Use machine learning algorithms to find more precise patterns for prediction (split data into a training and a test set, test set held out of training)

Example: n=50 ALFALFA



Example: Regression f_HI



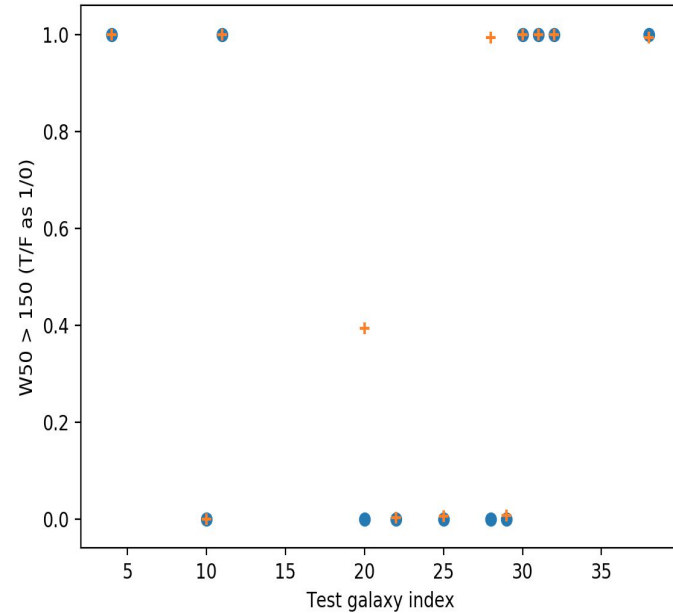
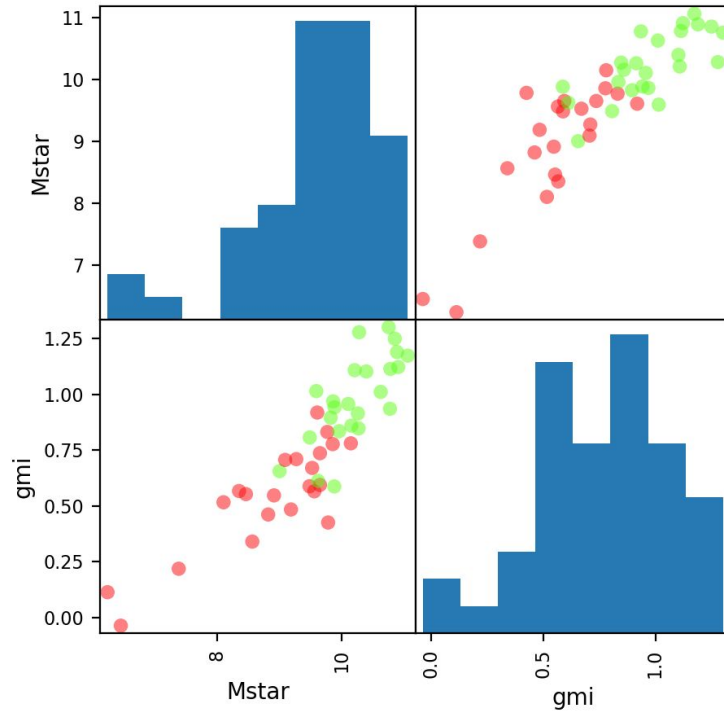
Example: Regression f_HI

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.model_selection import train_test_split

X = galaxies[['Mstar', 'gmi']]
y = galaxies['f_HI']
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)

est = GradientBoostingRegressor(n_estimators=100, max_depth=3)
est.fit(X_train, y_train)
pred = est.predict(X_test)
```

Example: Classification W50





Summer next steps

Obtain a full dataset (all objects, all wavelengths)

Include galaxy radii, light concentration, etc

Assess best machine learning algorithm to use

End product will predict f_{HI} or $\text{prob}(\text{det})$ without any prior knowledge of galaxy HI properties