STREAMLINE Training Summary Report: 2023-06-27 17:24:26.126216

General Pipeline Settings:

Data Path: /Users/yanbo/Dropbox/STREAMLINE-Regression/DemoData_2
Output Path:
/Users/yanbo/Dropbox/STREAMLINE-Regression/Colab_Output
Experiment Name: Demo_Experiment
Class Label: quality
Instance Label: None
Ignored Features: None
Specified Categorical Features: None
CV Partitions: 5
Partition Method: R
Match Label: None
Categorical Cutoff: 10
Statistical Significance Cutoff: 0.05
Export Feature Correlations: True
Export Univariate Plots: False
Random Seed: 42
Run From Jupyter Notebook: False
Use Data Scaling: True
Use Data Imputation: True
Use Multivariate Imputation: True
Use Mutual Information: True
Use MultiSURF: True
Use TURF: False
TURF Cutoff: 0.5
MultiSURF Instance Subset: 2000
Max Features to Keep: 2000
Filter Poor Features: False
Top Features to Display: 40
Export Feature Importance Plot: True
Overwrite CV Datasets: True
Primary Metric: explained_variance
Uniform Feature Importance Estimation (Models): True
Hyperparameter Sweep Number of Trials: 50
Hyperparameter Timeout: 900
Export Hyperparameter Sweep Plots: True
Export Metric Boxplots: True
Export Feature Importance Boxplots: True
Metric Weighting Composite FI Plots: explained_variance
Top Model Features To Display: 40

ML Modeling Algorithms:

Linear Regression: True
Elastic Net: True
Group Lasso: False
RF Regressor: False
AdaBoost: False
GradBoost: False
SVR: True

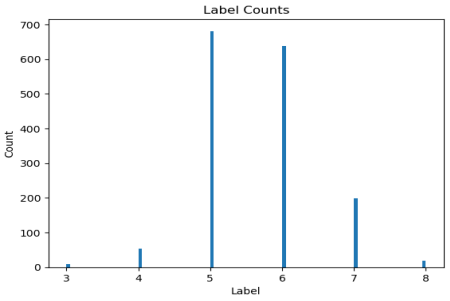
Datasets:

D1 = winequality-red

Univariate Analysis of Each Dataset (Top 10 Features for Each)

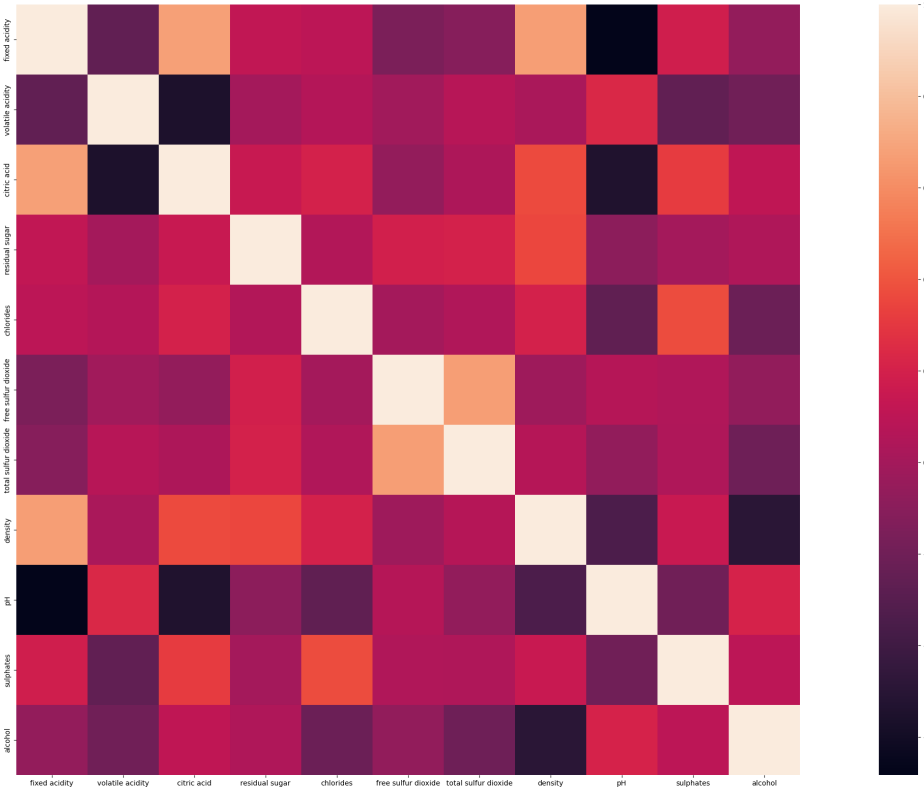
WARNING: Univariate analysis failed from scipy package error. To fix: pip install --upgrade scipy

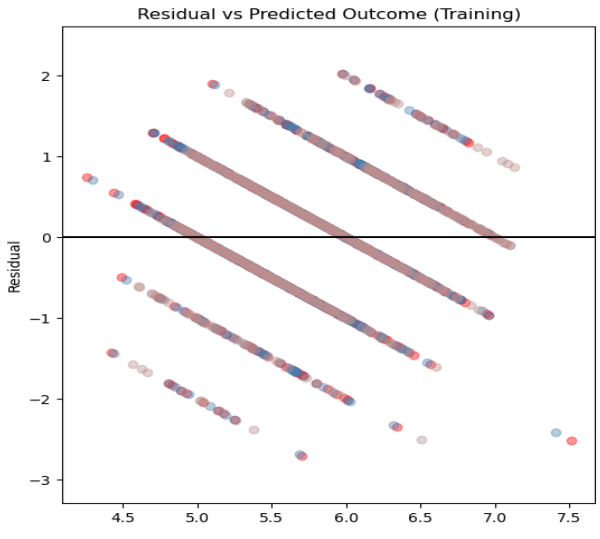
Dataset and Model Prediction Summary: D1 = winequality-red

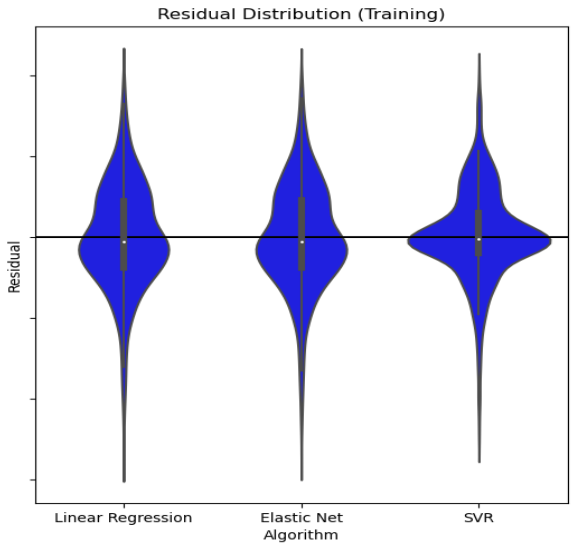


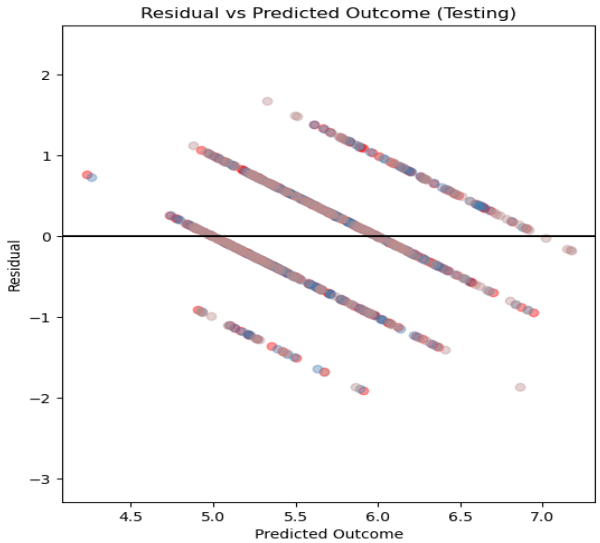
Dataset Counts Summary:
instances: 1599.0
features: 11.0
categorical_features: 0.0
quantitative_features: 11.0
missing_values: 0.0
missing_percent: 0.0

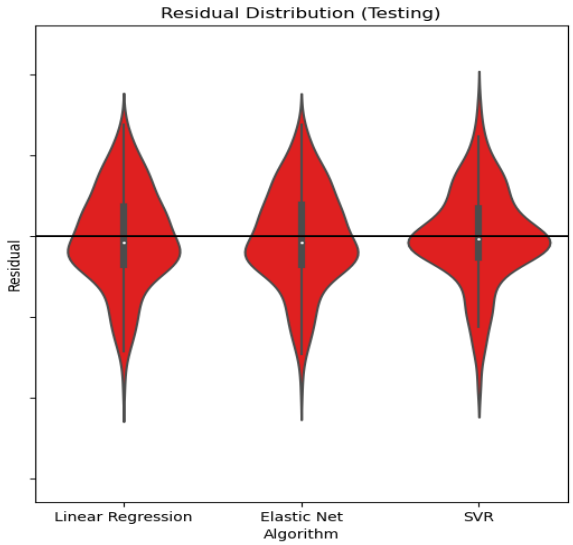
Top ML Algorithm Results (Averaged Over CV Runs):
Best (Max Error): Elastic Net = 2.229
Best (Mean Absolute Error): SVR = 0.461
Best (Mean Squared Error): SVR = 0.392
Best (Explained Variance): SVR = 0.398
Best (Median Absolute Error): SVR = 0.336
Best (Pearson Correlation): Elastic Net = 0.589





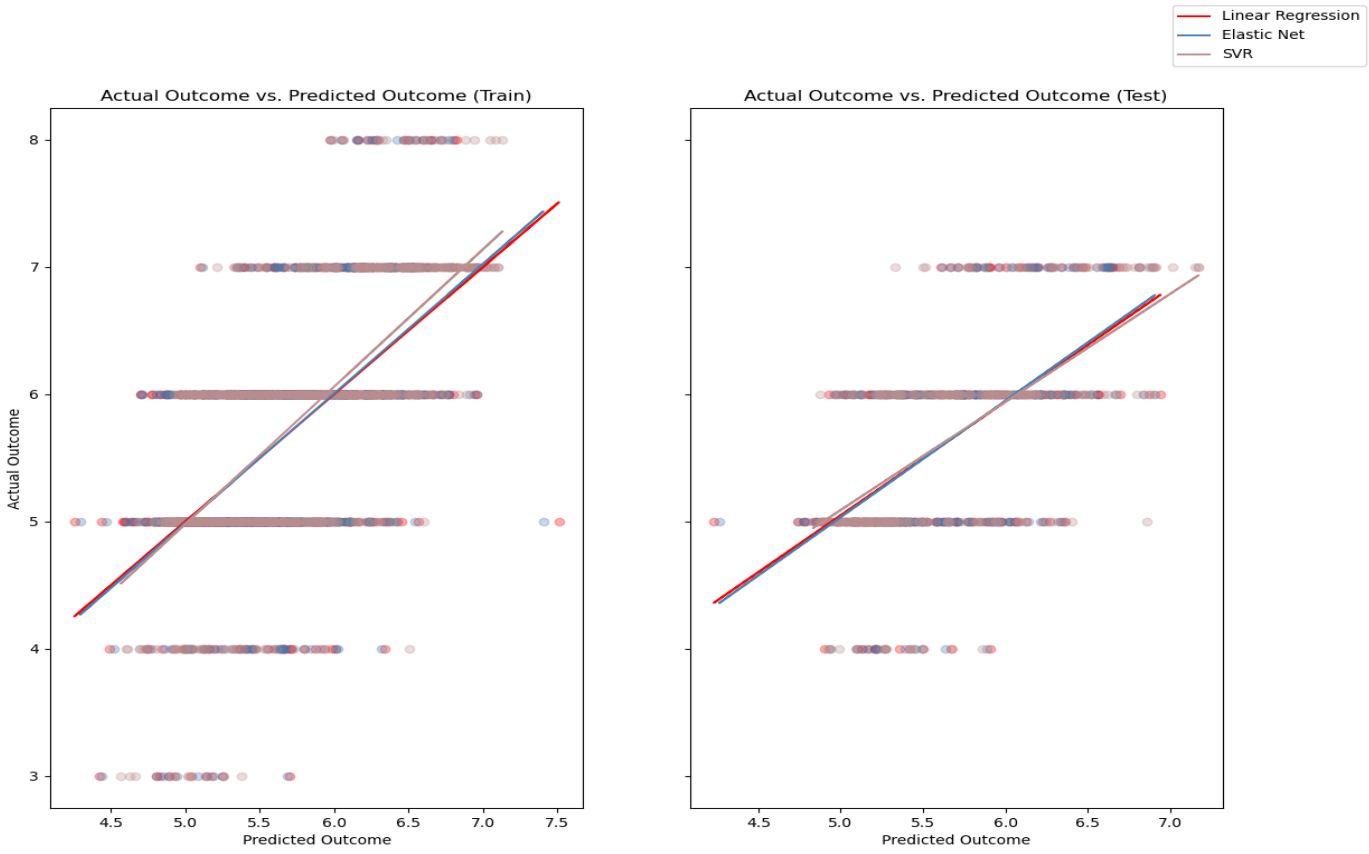


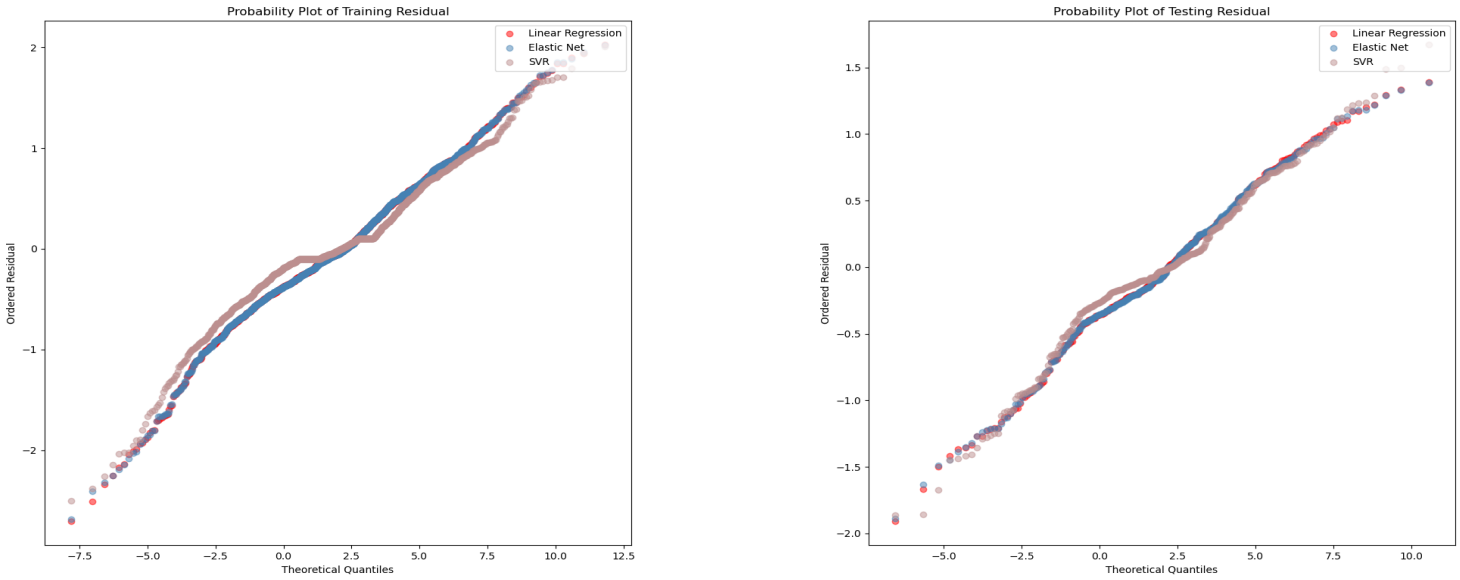




Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)

Dataset and Model Prediction Summary: D1 = winequality-red



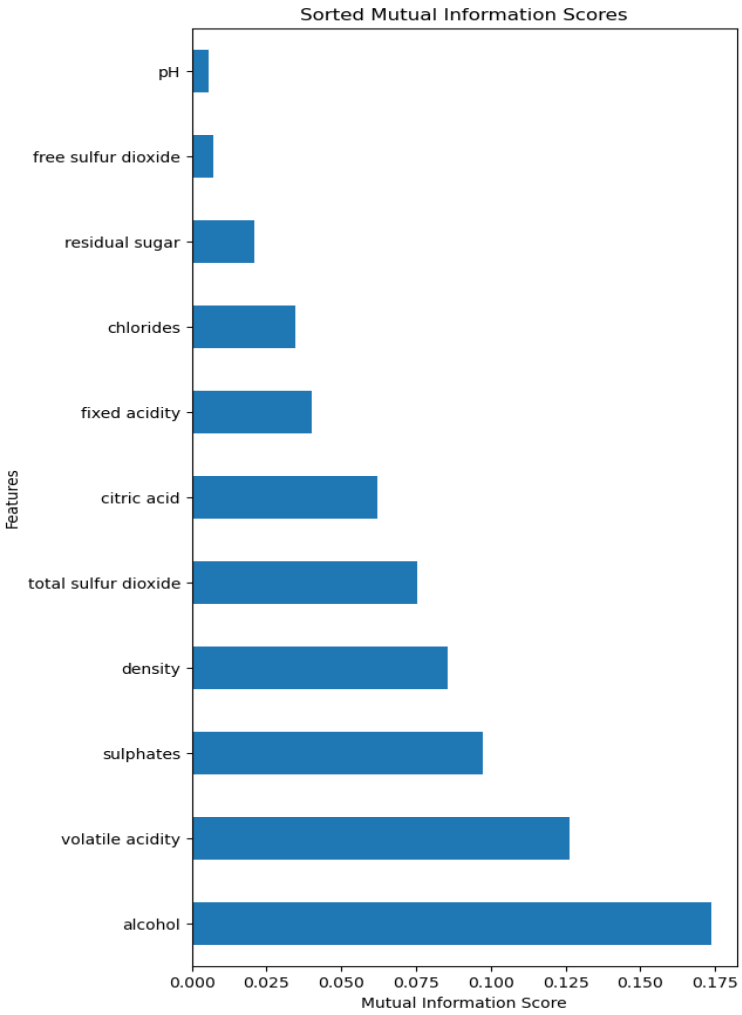


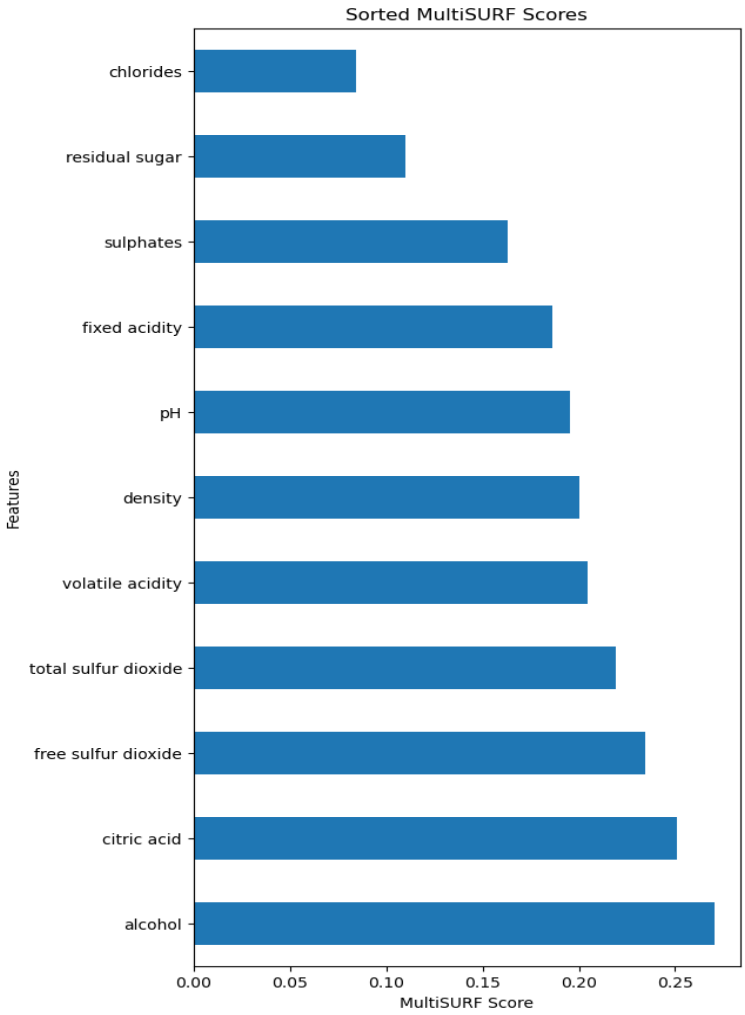
Average Model Prediction Statistics (Rounded to 3 Decimal Points)

D1 = winequality-red

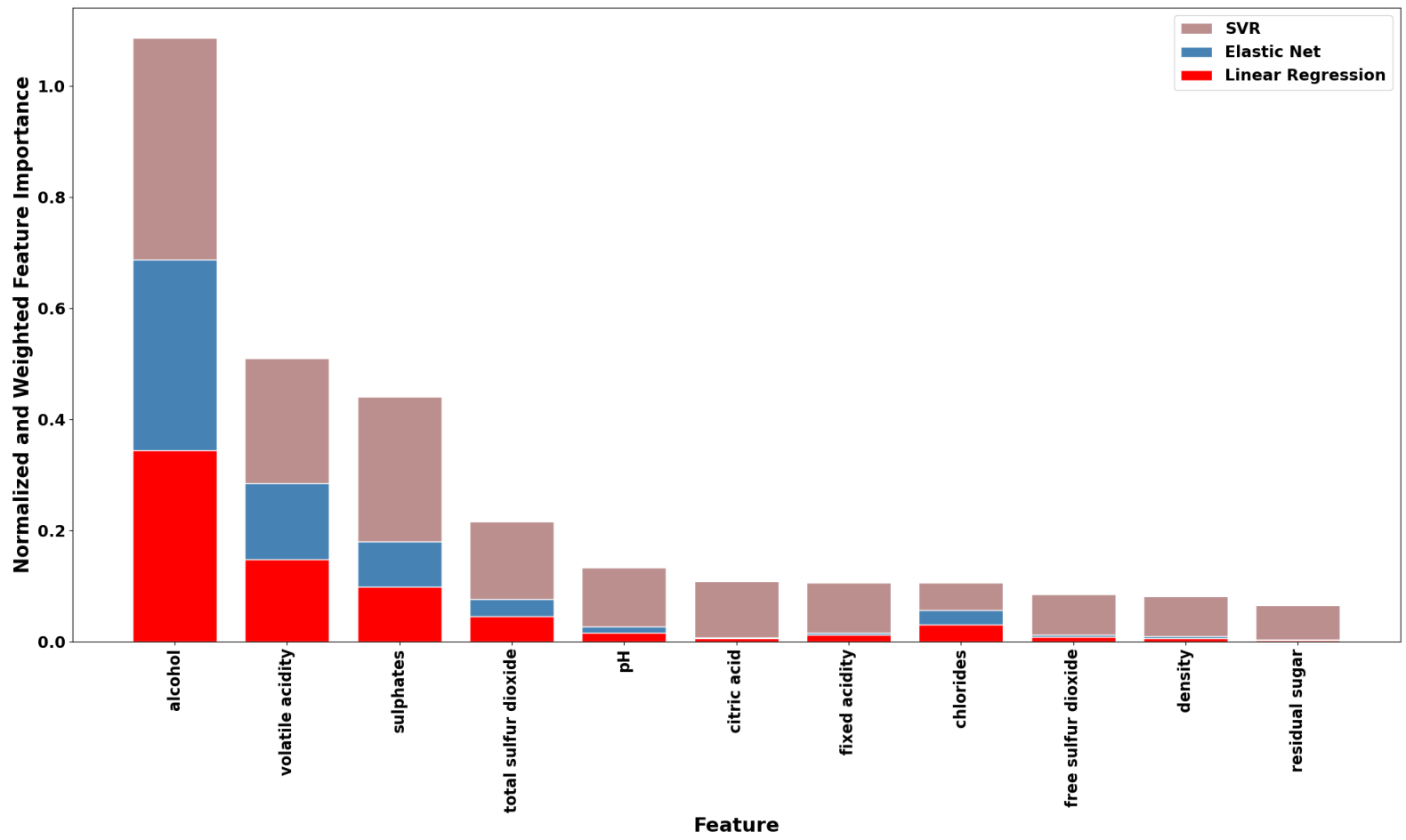
ML Algorithm	Max Error	Mean Absolute	Mean Squared	Median Absolute	Explained Variance	Pearson Correlation
Linear Regression	2.276	0.505	0.427	0.4	0.344	0.59
Elastic Net	2.229	0.508	0.428	0.407	0.343	0.589
SVR	2.455	0.461	0.392	0.336	0.398	0.634

Feature Importance Summary: D1 = winequality-red





Composite Feature Importance Plot (Normalized and Performance Weighted)



Generated with the URBS-Lab STREAMLINE: (https://github.com/UrbsLab/STREAMLINE)

Compare ML Performance Across Datasets

Using Best Performing Algorithms (Kruskall Wallis Compare Datasets)

Datasets:
D1 = winequality-red

Pipeline Runtime Summary

winequality-red

Pipeline Component	Time (sec)
Exploratory Analysis	0.79
Preprocessing	0.02
Mutual Information	0.34
MultiSURF	57.81
Feature Selection	0.29
Linear Regression	0.18
Elastic Net	12.34
SVR	124.58
Stats Summary	22.86