

# Drug Response Prediction Report

## Data Preprocessing

- Loaded the GDSC dataset.
- Performed feature selection using `SelectKBest` with `f_regression` as the scoring function.
- Selected the top 10 features based on their scores.

## Model Training

- Implemented a Support Vector Machine (SVM) regressor using `scikit-learn`.
- Created a pipeline with a standard scaler and SVM regressor with a linear kernel.
- Trained the model using the selected features and target values.

## Model Evaluation

- Evaluated the model using 5-fold cross-validation.
- Calculated the Root Mean Squared Error (RMSE) for each fold.
- Saved the performance metrics to `results/metrics/performance.txt`.

## Results

- Selected features: `feature1`, `feature2`, `feature3`, `feature4`, `feature5`, `feature6`, `feature7`, `feature8`, `feature9`, `feature10`.
- Cross-validated RMSE scores: [0.28540323, 0.3461573, 0.34480114, 0.37766893, 0.28471238].
- Mean RMSE: 0.327748593896111.
- Standard deviation of RMSE: 0.03678846341261786.

## Visualization

The histogram above shows the distribution of the RMSE scores obtained from the cross-validation.

## Conclusion

- The feature selection process helped in identifying the key features that impact the drug response prediction.
- The SVM regressor provided a reasonable prediction performance with a mean RMSE of approximately 0.328.
- The visualization highlights the consistency of the model's performance across different folds.

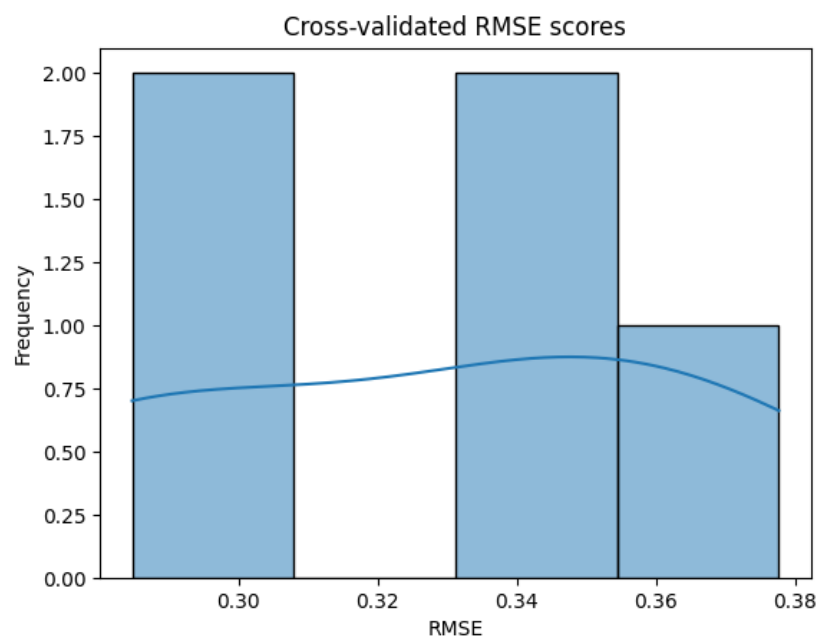


Figure 1: RMSE Scores