ICS 435 Final Project Report: An Exploratory Analysis of Powerlifting With Machine Learning

Hunter Von Tungeln, Lionel Roxas

April 29, 2025

0 GitHub Repository For Our Implementations

Our GitHub repository can be found here for specific implementation and instructions of how to run our models here: https://github.com/435FinalProject/OpenPowerliftingAnalysis.

1 Background and Problem

1.1 Background

Powerlifting is a sport consisting of three main barbell lifts/events, the squat, the bench press, and the deadlift. For each movement, there are three attempts to perform. The order is three attempts for squat, three attempts for bench press, then three attempts for deadlift. All three of these movements are judged by three judges, two on the sides, and one in the front. These lifts must be performed to a certain standard in order to be called a "good lift." Judges will either give a white light for a good lift, or a red light for a lift that is no-good. To be counted as a good lift, a lifter must receive at least two white lights from the judges. If a lifter receives at least one white-light, they may contest the judge's decisions. After the first attempt, lifters select weights for their second attempt, and after their second attempt, lifters select weights for their third attempt. Lifters are assigned to different weight classes and divisions based on their bodyweight and age. A lifter's performance is assessed by how much weight they squat, bench press, and deadlift. A lifter's total is the combined weight of their heaviest successful attempt for each lift.

1.2 Problem Statements

This project aims to answer a couple of questions. First, how well can machine learning models predict a lifter's age and weight based on their meet performance? Next, how well can machine learning models predict a lifter's best squat, bench press, and deadlift in a meet? Third, how well do these models perform on lifters of different levels (i.e. general powerlifting population vs. elite powerlifters)?

2 Motivation

2.1 Predicting Lifter Demographics (Age and Weight)

Powerlifting is a relatively small, niche sport with not very much money in it. The open-source dataset used in this project, which will be discussed later, OpenPowerlifting [10], relies on the powerlifting community to help and contribute to the maintainence records of all powerlifting competitions in history. Sometimes, competition data may be lost or corrupted. For example, a lifter's age or recorded weight may be lost. One way to mitigate this would be to provide a machine learning model's estimate of a lifter's demographics based on other pieces of data, like their squat or bench performance.

2.2 Predicting Lifter Performance

Regarding lifter performance, all attempts are important for a lifter to make in order to build their biggest possible total. Arguably, a lifter's third attempt is their most important because it is usually their heaviest lift for a given event, and a lifter is likely going to be attempting a weight very close to or above their personal best. Coaches/handlers (a handler is a meet-day coach that picks attempts and assists their lifter in navigating through a meet) need to have good attempt selection in order for their lifter to not miss/fail an attempt. These attempt selections are usually pre-planned based on a lifter's training and prior attempt outcomes. A machine learning based solution that helps predict a lifter's third attempt based on prior performances and factors (i.e. performance in the other two lifts, age, weight, etc.) would assist coaches and handlers greatly.

3 Solution and Methodology

We employ a supervised machine learning approach to develop predictive models for the above features. We use a filtered open-source dataset (OpenPowerlifting) with competition results of lifters across the two biggest powerlifting federations in the United States. This dataset contains lifter demographics of age and weight, as well as competition results. Competition results include a lifters successful and missed attempts. To identify the most accurate prediction method, we compare multiple regression models including both linear approaches (Ridge, Lasso, ElasticNet) and more complex tree-based ensembles (Random Forest and Gradient Boosting). The models are evaluated using standard regression metrics including R-squared, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) to determine which model is the most effective given our use cases. Hyperparameter tuning and cross-validation techniques are used in order for our models to generalize well to unseen data.

4 Related Work

Some work has been done in this area of predicting powerlifting performance and metrics given the OpenPowerlifting dataset. One group of researchers created a tool that allowed lifters to input their demographic data, lift information, equipment, and number of days until competition to predict how well a lifter would perform on meet day [21]. This group used similar models for these predictions, like Random Forest, Ridge Regression, and Lasso Regression for their predictions [21]. This is similar to what we are doing, with hopes of predicting Squat/Bench/Deadlift performance given our features. Two other papers propose new machine learning models/algorithms for predicting powerlifting performance based on a lifters attributes, like age, bodyweight, and best lifts to project and predict future performances [18], [20]. Another paper explores relationship between age and bodyweight on powerlifters, more specifically how age and weight correlate with performance in all three lifts [19]. Our approach is slightly different in the sense that we use regression models to predict bodyweight and age, instead of using them to predict performance. In this project, we use performance to predict bodyweight and age, and take note of the most important features to our model to see what correlates the most with those two demographics. Our approach is also different because we engineer more features into our dataset given existing features. All in all, powerlifting on its own is already a niche sport, hence why there is not much work being done at the intersection of powerlifting and machine learning.

5 Data and Preprocessing

5.1 Dataset Used

The dataset we used for this project comes from OpenPowerlifting [7], [10], a community-maintained archive of Powerlifting meet data. Their database is kept as a CSV file containing powerlifting meet records for people that compete. Each row represents one lifter at a given powerlifting meet. The columns, according to [7], are as follows:

Table 1a: OpenPowerlifting Dataset Columns and Description

Attribute	Description
Name	Lifter's Name
Sex	Lifter's Sex
Event	Type of Competition
Equipment	Type of Equipment Lifters Used
Age	Lifter's Age
AgeClass	Lifter's Age Class
BirthYearClass	Lifter's BirthYear Class
Division	Lifter's Competition Division
BodyweightKg	Lifter's Bodyweight in Kilograms
WeightClassKg	Lifter's WeightClass in Kilograms
Squat1Kg	Lifter's first squat attempt in Kilograms
Squat2Kg	Lifter's second squat attempt in Kilograms
Squat3Kg	Lifter's third squat attempt in Kilograms
Bench1Kg	Lifter's first bench press attempt in Kilograms
Bench2Kg	Lifter's second bench press attempt in Kilograms
Bench3Kg	Lifter's third bench press attempt in Kilograms
Deadlift1Kg	Lifter's first deadlift attempt in Kilograms
Deadlift2Kg	Lifter's second deadlift attempt in Kilograms
Deadlift3Kg	Lifter's third deadlift attempt in Kilograms

Table 1b: OpenPowerlifting Dataset Columns and Description

Attribute	Description
Squat4Kg	Lifter's fourth squat attempt (if granted) in Kilograms
Bench4Kg	Lifter's fourth bench-press attempt (if granted) in Kilograms
Deadlift4Kg	Lifter's fourth deadlift attempt (if granted) in Kilograms
Best3SquatKg	Lifter's maximum successful squat in kilograms
Best3BenchKg	Lifter's maximum successful bench press in kilograms
TotalKg	Sum of lifter's best successful squat, bench, and deadlift in kilograms
Place	Lifter's rank in their division/weight class
Dots	Formula to score powerlifters
Wilks	Formula to score powerlifters
Glossbrenner	Formula to score powerlifters
Goodlift	Formula to score powerlifters
Tested	Whether or not the federation the lifter competed in is drug tested
Country	Lifter's Home Country
State	Lifter's Home State/Province
Federation	Federation hosting the meet
ParentFederation	Parent of whatever federation is hosting the meet, if applicable
MeetCountry	Country of meet
MeetName	Name of meet
Sanctioned	If the meet was officially sanctioned by a recognized federation

Notes About Features That Are Important For Later

Sex Lifters are classified as M (Male), F (Female), or Mx (Non-Binary). We are only concered with male or female lifters, and will later separate the data into male and female sets.

Event Lifters can compete in different types of events, most commonly being SBD (Squat-Bench-Deadlift) or "Full-Power". Lifters can also compete BD (Bench-Deadlift), SD (Squat-Deadlift), SB (Squat-Bench), Squat only, Bench Only, and Deadlift only. We are concerned with lifters that compete in full-power meets, i.e. only SBD.

Equipment Lifters can choose different types of equipment to compete with. Raw means they are only allowed neoprene knee sleeves. Wraps mean they are allowed knee wraps. Single-ply means they are allowed single-ply suits to assist their lifting. Multi-ply means they are allowed multi-ply suits for assistance. Unlimited essentially means anything supportive is allowed. Straps means a lifter is allowed to use straps for deadlift. We are concerned with Raw powerlifting for this project, as it is currently the most popular form of the sport.

Country/State We only look at powerlifting in the United States for this project.

Federation Furthermore, we only look at the two most popular powerlifting federations in the United States, being USA Powerlifting (USAPL), and Powerlifting America (AMP).

Wilks, Dots, Glossbrenner, Goodlift These are all different formulas for scoring powerlifting performance to enable comparison of lifters across different weight classes. The exact formulas are not important for this analysis.

Tested We are only concerned with lifters that compete in a drug-tested federation (both USAPL and AMP are tested).

Sanctioned We are only concerned with official, sanctioned meets.

Squat/Bench/Deadlift4KG These rarely occur in meets, only due to an official or technical problem, so these are not important for this analysis.

Squat/Bench/Deadlift Positive values represent successful attempts, negative values represent missed lifts. We handle this below in preprocessing.

5.2 Preprocessing

5.2.1 Preliminary Filtering

We first filter the dataset by filtering out competitors that did not compete as "full-power" lifters and that are not "Raw" lifters for equipment. We filter out untested lifters, as we are more interested in what the "normal" human can do while not on any performance enhancing drugs. To control the size of our data, we only look at the two most popular powerlifting federations in the United States, USAPL and AMP, and filter out the rest. We filter out any meet that is not sanctioned, as we are only interested in official results. We filter out any lifter with a NaN TotalKg, as that indicates they did not complete at least one successful squat, bench, and deadlift. Then, we eliminate columns that have no relation to a lifter's performance, like where the meet took place, what state the meet was in, their if the meet was sanctioned, etc. because these columns have no relation to how a lifter will perform. The columns dropped were Squat4Kg, Bench4Kg, Deadlift4Kg, AgeClass, BirthYearClass, MeetCountry, MeetState, MeetTown, MeetName, Sanctioned, ParentFederation, Federation, Tested, Country, State, Date, Place, WeightClassKg, Divison. Obviously, WeightClassKg and AgeClass must be dropped, as our tasks consist of predicting age and weight based on powerlifting performance.

5.2.2 Feature Engineering

We then add some new columns based on existing data present. We add:

1. Lift Ratios: Proportion of each lift to total (i.e. squat/total ratio).
2. Strength to bodyweight ratios: Proportion of each lift to bodyweight (i.e. bench/bodyweight ratio).
3. Binary Encoding of sex: Male and female columns with a 1 representing yes and 0 representing no.
4. Successful lift indicators: Additional columns indicating whether or not a given attempt in a meet was successful, 1 representing successful, 0 representing unsuccessful.

Then, we fill NaN values for attempts with 0, since we now have indicators as to whether or not a lift was successful.

5.2.3 Finalizing the Datasets

Now, we drop the Name, Sex, Event, and Equipment columns, as well as drop any rows with residual NaN values, and partition the dataset into two parts, male and female. The datasets are now ready for use with machine learning models. Our final list of features for each dataset is as follows:

Table 2a: Preprocessed OpenPowerlifting Dataset Columns and Description

Attribute	Description
Age	Lifter's Age
BodyweightKg	Lifter's Bodyweight in Kilograms
Squat1Kg	Lifter's first squat attempt in Kilograms
Squat2Kg	Lifter's second squat attempt in Kilograms
Squat3Kg	Lifter's third squat attempt in Kilograms
Best3SquatKg	Lifter's best squat from attempts
Bench1Kg	Lifter's first bench press attempt in Kilograms
Bench2Kg	Lifter's second bench press attempt in Kilograms
Bench3Kg	Lifter's third bench press attempt in Kilograms
Best3BenchKg	Lifter's best bench press from attempts
Deadlift1Kg	Lifter's first deadlift attempt in Kilograms
Deadlift2Kg	Lifter's second deadlift attempt in Kilograms
Deadlift3Kg	Lifter's third deadlift attempt in Kilograms
Best3DeadliftKg	Lifter's best deadlift from attempts
TotalKg	Total of best squat, bench, and deadlift

Table 2b: Preprocessed OpenPowerlifting Dataset Columns and Description

Attribute	Description
Dots	Powerlifting scoring formula (Dots)
Wilks	Powerlifting scoring formula (Wilks)
Glossbrenner	Powerlifting scoring formula (Glossbrenner)
Goodlift	Powerlifting scoring formula (Goodlift)
Squat1Success	Success or fail on first squat attempt
Squat2Success	Success or fail on second squat attempt
Squat3Success	Success or fail on third squat attempt
Bench1Success	Success or fail on first bench attempt
Bench2Success	Success or fail on second bench attempt
Bench3Success	Success or fail on third bench attempt
Deadlift1Success	Success or fail on first deadlift attempt
Deadlift2Success	Success or fail on second deadlift attempt
Deadlift3Success	Success or fail on third deadlift attempt
Squat_BW_Ratio	Best squat divided by bodyweight
Bench_BW_Ratio	Best bench press divided by bodyweight
Deadlift_BW_Ratio	Best deadlift divided by bodyweight
Squat_Total_Ratio	Best squat divided by total
Bench_Total_Ratio	Best bench press divided by total
Deadlift_Total_Ratio	Best deadlift divided by total

For each task, we drop the associated column and use the rest of the columns as features to predict on. For example, if we wanted to predict bodyweight, we drop the bodyweight column and use it as our target variable, with the rest of the features being used by our models for predictions. We also scale the data using Scikit-Learn's StandardScaler() function, before giving the data to the models.

6 Model Selection, Hyperparameters and Performance Metrics

6.1 Models Used

We chose six different regression models for these tasks. The first model, a Decision Tree Regressor, is used in order to tune hyperparameters that will be used in other tree based models. The next five are trained end evaluated on a test set. The five models being evaluated are a Decision Tree Regressor, a Random Forest Regressor, a Gradient Boosted Decision Tree Regressor, ElasticNet Regression, Ridge Regression, and Lasso Regression.

6.1.1 Decision Tree Regressor

The first model, we chose was the Decision Tree Regressor [3]. This model is very simple in that it chooses the best "split" on data based on a given criterion, like mean-sqaured error as we used. The model splits data into multiple subsets, and keeps splitting until there is no data left to split on. The model predicts a continuous numerical value based on a given set of features and real values. We did not use this model for training and testing results, rather, we used this model to tune hyperparameters for the Random Forest Regressor and Gradient Boosted Decision Tree Regressor, as both of those are tree based models. Tuning hyperparameters on a regular decision tree gives some insight as to what hyperparameters to use in our other tree based models.

6.1.2 Random Forest Regressor

The first model we chose to evaluate our results on is the Random Forest Regressor implemented in the Scikit-Learn library [11]. This model is an ensemble of decision tree regressors, that fit copies of each tree on subsets of a given dataset, averaging their prediction to predict a final numerical value.

6.1.3 Gradient Boosted Decision Tree Regressor

The next model, the Gradient Boosted Decision Tree Regressor, implemented in the Scikit-Learn library, is an ensemble of trees, similar to a Random Forest Regressor. The model builds trees in an iterative fashion, with each subsequent tree trying to correct the errors of past trees. The goal of this iterative approach is to build the most accurate decision tree possible, while minimizing a chosen loss function. See [5] for more information.

6.1.4 ElasticNet Regression

The third type of model we are evaluating is ElasticNet Regression . ElasticNet combines both L1 and L2 penalties, and is essentially a hybrid of both Ridge and Lasso regression, by linearly combining the two, with an intermediate hyperparameter that controls the balance between both L1 and L2 loss. ElasticNet also seeks to minimize the mean squared error between actual and predicted values. See [4] for more information.

6.1.5 Ridge Regression

The fourth type of model we are evaluating is a Ridge Regression model. Ridge Regression (L2) is a modified ordinary least squares model that penalizes high feature weights with a regularization term, and seeks to minimize the mean squared error between actual and predicted values. See [16] and [12] for more information.

6.1.6 Lasso Regression

The last type of model we evaluate is a Lasso Regression model. Lasso regression (L1) multiplies a penalty term by a regularization parameter, then adds the scaled penalty term to the residual sum of squares, with the goal of minimizing the mean squared error. See [14] and [8] for more information.

6.2 Hyperparameter Optimization

Hunter All of the above models were evaluated using with a random_state = 42 parameter for reproducibility. All model hyperparameters were tuned using Scikit-Learn's GridSearchCV function, which systematically tests each combination of hyperparameters for a given model with cross-fold validation to find the best performing model (see [6] for an in-depth explanation). GridSearchCV was used with the parameters 'cv' = 5 and 'scoring' ='neg_mean_squared_error'. 'cv' controls how many folds are used for cross-validation, which evaluates model performance on different subsets of unseen data in order to make sure the model generalizes well. See [1] for more information regarding cross-validation. We used mean squared error because our problem is a regression problem that seeks to minimize mean squared error.

6.2.1 Decision Tree Regressor

As stated before, a Decision Tree Regressor was used in order to leverage insights on what hyperparameters to use for our other tree based models. The three main hyperparameters we tuned for this model were max_depth, min_samples_split, and min_samples_leaf. The first hyperparameter, max_depth, controls how deep the decision tree splits. The second hyperparameter, min_samples_split controls how many samples need to be considered in order for a split to occur. The third hyperparameter, min_samples_leaf controls how many samples are required for a node to be considered a leaf, i.e. no more splitting can occur at that node. See [11] for more information. The sets of parameters tested and tuned included 'max_depth': (None, 1, 3, 5, 10, 20, 30), 'min_samples_split': (2, 5, 10, 20, 50), and 'min_samples_leaf': (1, 2, 5, 10, 15, 25). Other notable parameters that were kept constant for this model were the criterion, which 'friedman_mse' was used, and 'splitter = best'. All other parameters were kept as the default as of Scikit-Learn version 1.6.1.

6.2.2 Random Forest Regressor

For the Random Forest Regressor, we only tuned one hyperparameter, that being 'n_estimators', which essentially the amount of trees included in the ensemble. Other hyperparamters that were hand picked based on the Decision Tree Regressor results were max_depth, min_samples_split, and min_samples_leaf. We also used 'boostrap = True', which essentially allows each tree to sample the same rows of data when training. All other parameters were kept as the default as of Scikit-Learn version 1.6.1. See [11] for more details on each hyperparameter.

6.2.3 Gradient Boosted Decision Tree Regressor

For the Gradient Boosted Decision Tree Regressor, we tuned 'learning_rate', and kept 'n_estimators' = 100. Again, other hyperparamters that were hand picked based on the Decision Tree Regressor results were max_depth, min_samples_split, and min_samples_leaf. The learning rate controls the contribution of each tree to the entire ensemble. Loss was kept as the default of squared error. All other parameters were kept as the default as of Scikit-Learn version 1.6.1. See [5] for more details on each hyperparameter.

6.2.4 ElasticNet Regression

For the ElasticNet Regression models, we tuned 'alpha' and 'L1_ratio'. The 'alpha' parameter controls the constant that is multiplied by the penalty terms, essentially controlling how strong regularization of the model is. The 'L1_ratio' controls the balance between L1 and L2 loss used by the model. The sets of hyperparameter values included 'alpha': (0.1, 1.0, 10.0), and 'L1_ratio': (0.1, 0.5, 0.7, 1.0). All other parameters were kept as the default as of Scikit-Learn version 1.6.1. See [4] for more information on each hyperparameter.

6.2.5 Ridge Regression

For the Ridge Regression models, we tuned the 'alpha' parameter, which controls the penalty or regularization of the model by penalizing high feature weights to prevent overfitting. The included values of 'alpha' that we tuned are (0.01, 0.1, 1.0, 10.0). For this model, we used 'max_iter' = 10000, and 'solver = 'saga''. The 'max_iter' parameter controls how many iterations are allowed in order for the model to reach convergence, and 'solver' controls the method used for optimization. The 'saga' solver uses Stochastic Average Gradient Descent, which seeks to minimize the gradient of our loss function to improve model performance. See [12] for more information on each parameter, and see [9] for more information on Stochastic Gradient Descent.

6.2.6 Lasso Regression

For the Lasso Regression models, we tuned the 'alpha' parameter, which, as explained above, similar to Ridge Regression, controls the penalty or regularization of the model by penalizing high feature weights to prevent overfitting. The included values of 'alpha' that we tuned are (0.01, 0.1, 1.0, 10.0). For this model, we used 'max_iter' = 50000 to ensure the model converged to a minimized loss. See [8] for more information on each parameter.

6.3 Performance Metrics

Three metrics were used to evaluate performance and interpretations of our models, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the R^2 value.

6.3.1 Root Mean Squared Error (RMSE)

Root mean squared error is represented as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} |y(i) - \hat{y}(i)|^2}{N}}$$

Where $y(i)$ is the actual value, $\hat{y}(i)$ is the predicted value, and N is the total number of samples. The top half of the fraction is the sum of squared differences between the actual and predicted values, which then gets divided by the number of samples. The square root of that quotient is taken to obtain the RMSE. The RMSE will essentially tell us how close our model is to predicting a value, so a lower RMSE is better. See [2] , [13], [17] for more information on this error metric.

6.3.2 Mean Absolute Error (MAE)

Mean Absolute Error, or MAE, is simply the sum of differences between the predicted and actual value divided by the total number of samples. Mathematically, this is represented as:

$$MAE = \frac{\sum_{i=1}^{n} |y(i) - \hat{y}(i)|}{N}$$

Like RMSE, a lower MAE is better. However, compared to RMSE, MAE is more robust to outliers, but does not penalize them as much. We care more about accuracy of our predictions, so our emphasis is more on RMSE. See [2] and [15] for more information on this error metric.

6.3.3 R^2 Value

The R^2 value is known as the coefficient of determination (cite). Mathematically, it is represented as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y(i) - \hat{y}(i))^2}{\sum_{i=1}^{n}(y(i) - \bar{y})^2}$$

Where $y(i)$ is the actual value, $\hat{y}(i)$ is the predicted value, $\bar{y}$ is the mean of actual values. The R^2 value essentially tells us how much our features in our model correlate with our target values. This is more useful for drawing insights as to how our features relate to a predicted variable. See [2] for more information on this error metric.

7 Model Performance on Different Problems

Here we focus on the most important features and model performance for our different regression tasks.

7.1 Predicting Male Age

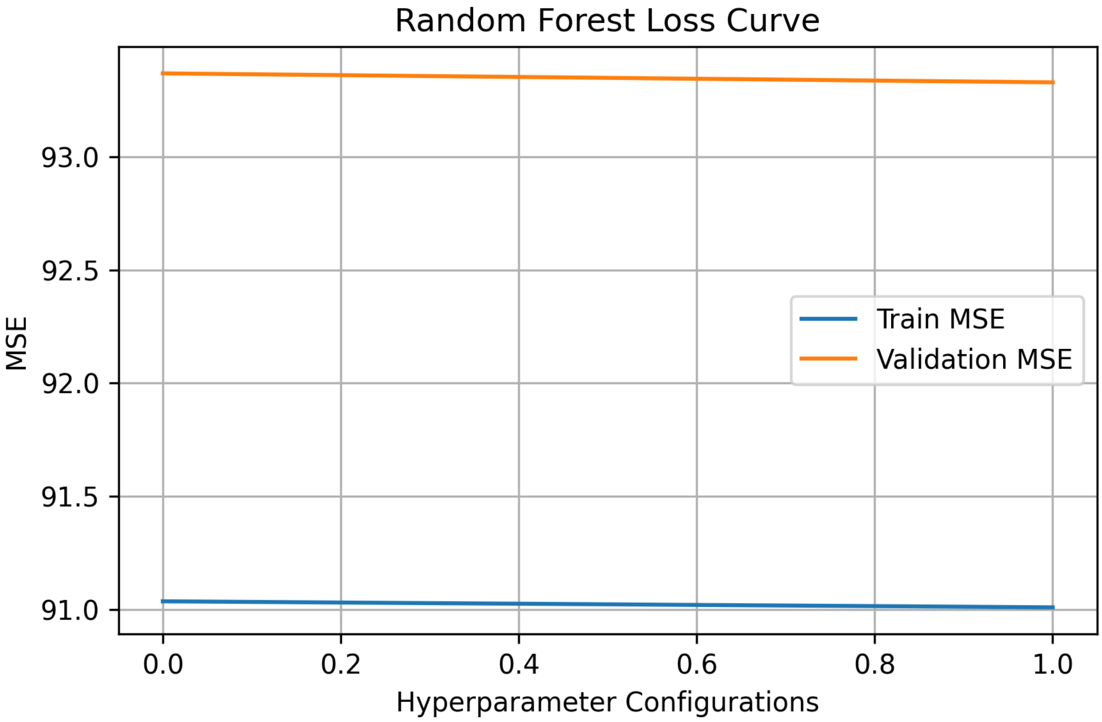
This section focuses on how well our models predicted age for male powerlifters.

7.1.1 Decision Tree Regressor

This model was used for tuning and gaining insights on hyperparameters to use on for our Random Forest Regressor and Gradient Boosted Decision Tree Regressor. The optimal hyperparameters found here during the use a GridSearchCV were a max_depth of 5, a min_samples_leaf of 25, and a min_samples_split of 2.

7.1.2 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, a min_samples_leaf of 25, a min_samples_split of 2, and an n_estimators of 200. The loss curve for our model is shown below:



Observe here that we only had 2 hyperparameter configurations, configuration 0, and configuration 1, where configuration 1 is our optimal model. We can observe from this plot that our MSE is very high on our validation sets, which is terrible given the task at hand.

The performance of our regression model on the test set is shown in a table below:

Table 3: Random Forest Regressor Model Performance for Predicting Male Age

Model	RMSE	MAE	R^2
Random Forest Regressor	9.615	6.641	0.160

In addition, we show the 10 most important and influential features for this model.

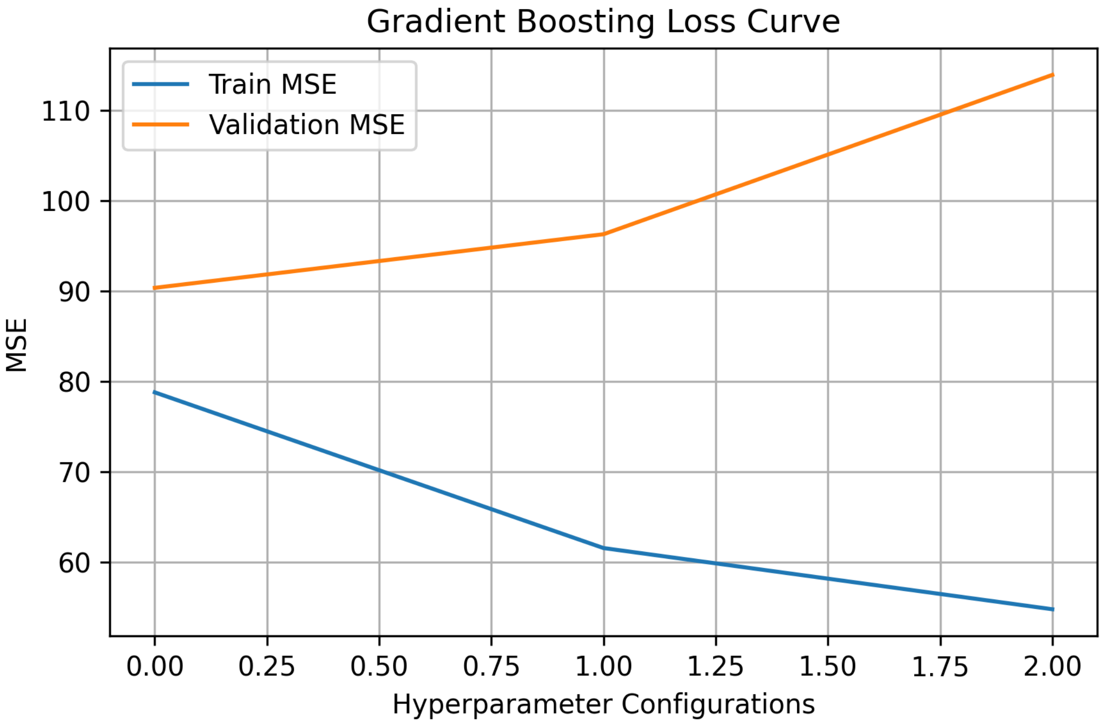
Table 4: Random Forest Regressor Feature Importance for Predicting Male Age

Feature Importance Ranking	Feature Name	Coefficient
1	Bench_Total Ratio	0.316
2	Bodyweight_Kg	0.268
3	Squat_BW_Ratio	0.225
4	Squat_Total_Ratio	0.086
5	Best3SquatKg	0.020
6	Best3BenchKg	0.016
7	Best3DeadliftKg	0.013
8	Deadlift_BW_Ratio	0.013
9	TotalKg	0.005
10	Goodlift	0.004

Interpretation: Our two error metrics, RMSE and MAE, are very high, meaning our model can predict someone's age given all of our features to be within about ± 6 years, with some bad predictions being within ± 10 years. This tells us that there are outliers present that siginificantly affect our RMSE. An optimal value would be a prediction range of 0.5-2 years. Our features have low coefficients, and the R^2 score is very low, meaning our features have almost no correlation with someone's age.

7.1.3 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, a min_samples_leaf of 25, a min_samples_split of 2, an n_estimators of 100, and a learning_rate of 0.1. The loss curve for our model is shown below:



Observe here that we only had 3 hyperparameter configurations, where the optimal model had the lowest MSE on the validation sets during cross-validation. We can observe from this plot that our MSE is again very high on our validation sets, as well as our training sets, which is terrible given the task at hand.

The performance of our regression model on the test set is shown in a table below:

Table 5: Gradient Boosted Decision Tree Regressor Model Performance

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	9.456	6.490	0.188

In addition, we show the 10 most important and influential features for this model.

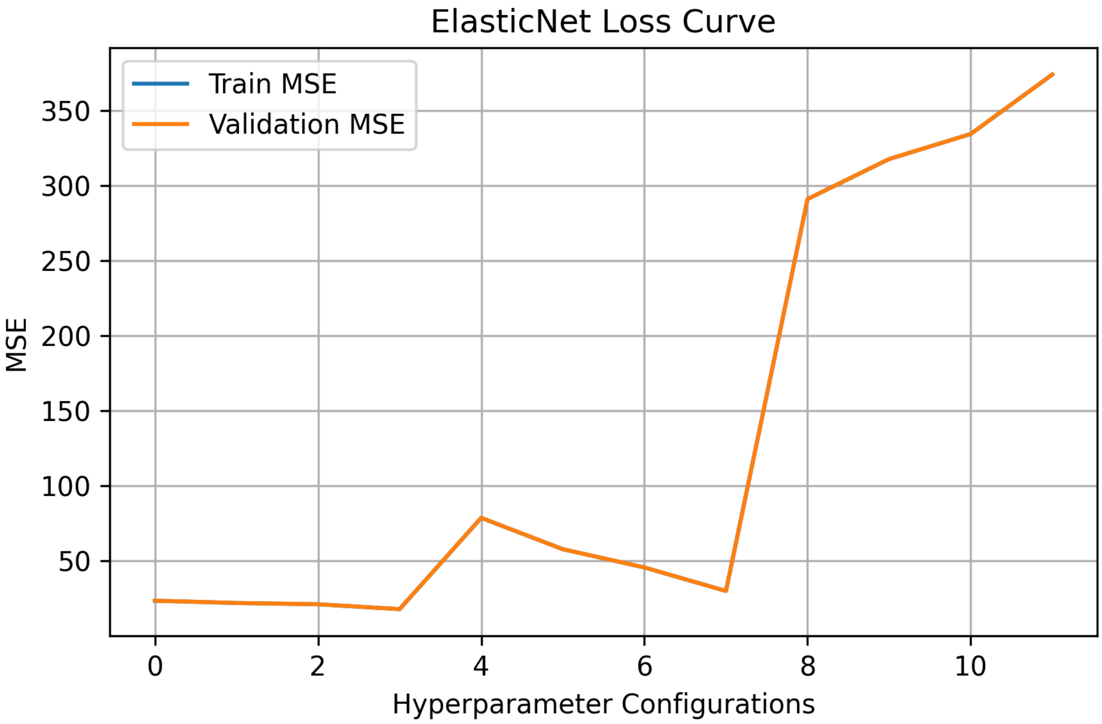
Table 6: Gradient Boosted Decision Tree Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Bench_Total_Ratio	0.2140
2	BodyweightKg	0.1811
3	Squat_BW_Ratio	0.1629
4	Squat_Total_Ratio	0.1097
5	Best3DeadliftKg	0.0378
6	Deadlift_BW_Ratio	0.0359
7	Bench_BW_Ratio	0.0236
8	Glossbrenner	0.0231
9	Deadlift_Total_Ratio	0.0200
10	Bench1Kg	0.0186

Interpretation: Again, like Random Forest, our two error metrics, RMSE and MAE, are fairly high, meaning at best, our model can predict someone's age given all of our features to be within ± 6 years, with some bad predictions being within ± 10 years. Interestingly, our MAE is lower than our RMSE by about 3, meaning there are a outliers that siginificantly affect our RMSE. Again, an optimal value would be a prediction range of around 0.5-2 years. Our features have low coefficients, and the R^2 score is very low, meaning our features have almost no correlation with someone's age.

7.1.4 ElasticNet Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.1. The loss curve for our model is shown below:



Observe here that we had 12 hyperparameter configurations. The best hyperparameter configuration above had the smallest MSE. However, we can observe from this plot that our MSE is very high on our validation sets, which is, again, terrible.

The performance of our regression model on the test set is shown in a table below:

Table 7: ElasticNet Model Performance

Model	RMSE	MAE	R^2
ElasticNet	9.831	6.815	0.122

We show the 10 most important and influential features for this model below.

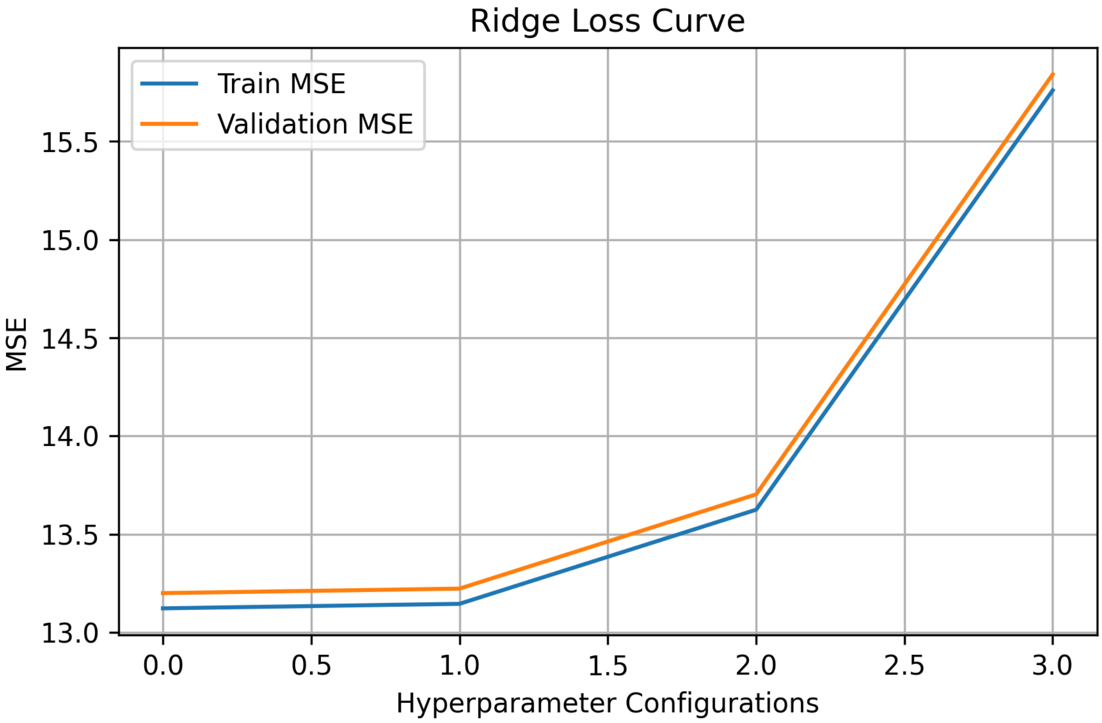
Table 8: ElasticNet Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Bench_Total_Ratio	1.4949
2	Squat_Total_Ratio	1.3969
3	BodyweightKg	1.3949
4	Glossbrenner	0.6135
5	Best3DeadliftKg	0.5452
6	Wilks	0.5408
7	Bench2Success	0.4231
8	Best3BenchKg	0.3831
9	Deadlift1Kg	0.3804
10	Bench3Success	0.3560

Interpretation: Like our previous two models, Our two error metrics, RMSE and MAE, are very high, meaning our model can predict someone's age given all of our features to be within about ±6 years, with some bad predictions being within ±10 years. This tells us that there are outliers present that siginificantly affect our RMSE. An optimal value would be a prediction range of 0.5-2 years. Our features have low coefficients, and the R^2 score is very low, meaning our features have almost no correlation with someone's age.

7.1.5 Ridge Regression

This model's optimal hyperparameters we foiund during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



There were 4 different hyperparameter configurations, with the best one having the hyperparameters listed above. Again, the MSE for each model on the cross-fold validation sets were high, but much less compared to the three previous models. The performance of our regression model on the test set is shown in a table below:

Table 9: Ridge Regression Model Performance

Model	RMSE	MAE	R^2
Ridge	9.750	6.730	0.136

We show the 10 most important and influential features for this model below.

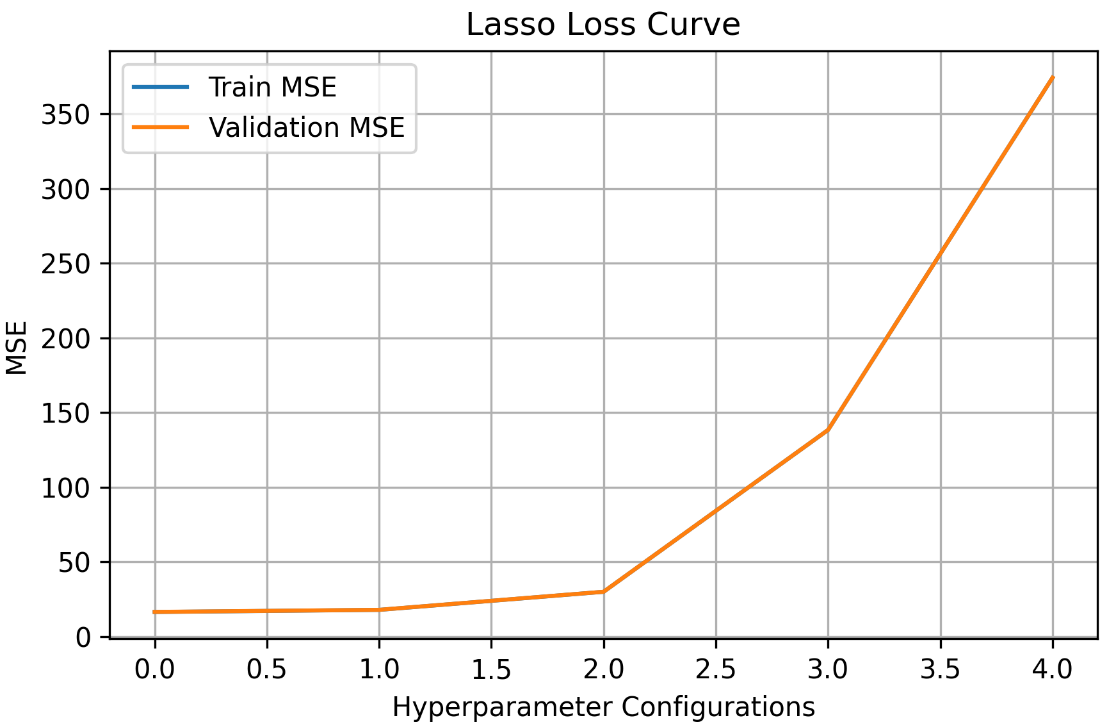
Table 10: Ridge Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Wilks	35.8046
2	Dots	26.6115
3	Glossbrenner	8.7806
4	Goodlift	6.9004
5	BodyweightKg	4.8835
6	Squat_BW_Ratio	4.8652
7	Squat_Total_Ratio	4.1631
8	Deadlift_BW_Ratio	3.4399
9	Bench_BW_Ratio	1.9425
10	Deadlift_Total_Ratio	1.6460

Interpretation: Similar to our previous models, Our two error metrics, RMSE and MAE, are very high, meaning our model can predict someone's age given all of our features to be within about ±6 years, with some bad predictions being within ±10 years. An interesting thing to note here is that Wilks, DOTS, Glossbrenner, and Goodlift were important features to this model, with Wilks and DOTS having very high coefficients. This is interesting as both Wilks and DOTS, as explained before, are 'score' formulas for powerlifting, which enable the comparison of lifters across weight classes. Other features were also given more emphasis compared to prior models, but this did not affect the outcome of our model performing poorly.

7.1.6 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



There were 5 different hyperparameter configurations, with the best one having the hyperparameters listed above. Again, the MSE for each model on the cross-fold validation sets, especially for this model, is very high. The performance of our regression model on the test set is shown in a table below:

Table 11: Lasso Regression Model Performance

Model	RMSE	MAE	R^2
Lasso	9.769	6.746	0.133

We show the 10 most important and influential features for this model below.

Table 12: Lasso Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Wilks	6.4864
2	Squat_BW_Ratio	4.1688
3	BodyweightKg	3.7250
4	Squat_Total_Ratio	2.3418
5	Bench_Total_Ratio	2.0992
6	Best3DeadliftKg	1.1010
7	Deadlift3Success	0.9988
8	Deadlift3Kg	0.9717
9	Deadlift_BW_Ratio	0.8603
10	Bench2Success	0.8332

Interpretation: Again, similar to our previous two models, Our two error metrics, RMSE and MAE, are very high, meaning our model can predict someone's age given all of our features to be within about ±6 years, with some bad predictions being within ±10 years. Again, an interesting thing to note here is that Wilks, was an important features to this model. However, this model still performed poorly on this task.

7.1.7 Comparison and Domain Interpretation

Table 13: Comparison of Model Performance for Predicting Male Age

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	9.456	6.490	0.188
Random Forest Regressor	9.615	6.641	0.160
Ridge	9.750	6.730	0.136
Lasso	9.769	6.746	0.133
ElasticNet	9.831	6.815	0.122

From the table above, we can observe that none of our above models performed well at predicting someone's age based on their powerlifting performances. However, this is to be expected, as the sport of powerlifting is open to everyone, regardless of strength, age, race, sex, etc. Anyone from any background can participate in a local powerlifting meet. Some people start off stronger than others, at a younger age, with some older people being significantly stronger than younger people. There are powerlifters that started at age 40 or older. At some meets that one of the authors (Hunter) attended, there were a couple of powerlifters in their 50s and 60s, that outlift some younger people. In addition, genetics play a huge role in strength at a given age, but that is beyond the scope of this paper and author knowledge. For these reasons, poor model performance when predicting age based solely on powerlifting attributes is to be expected, as these attributes alone do not correlate with age well, if at all.

7.2 Predicting Female Age

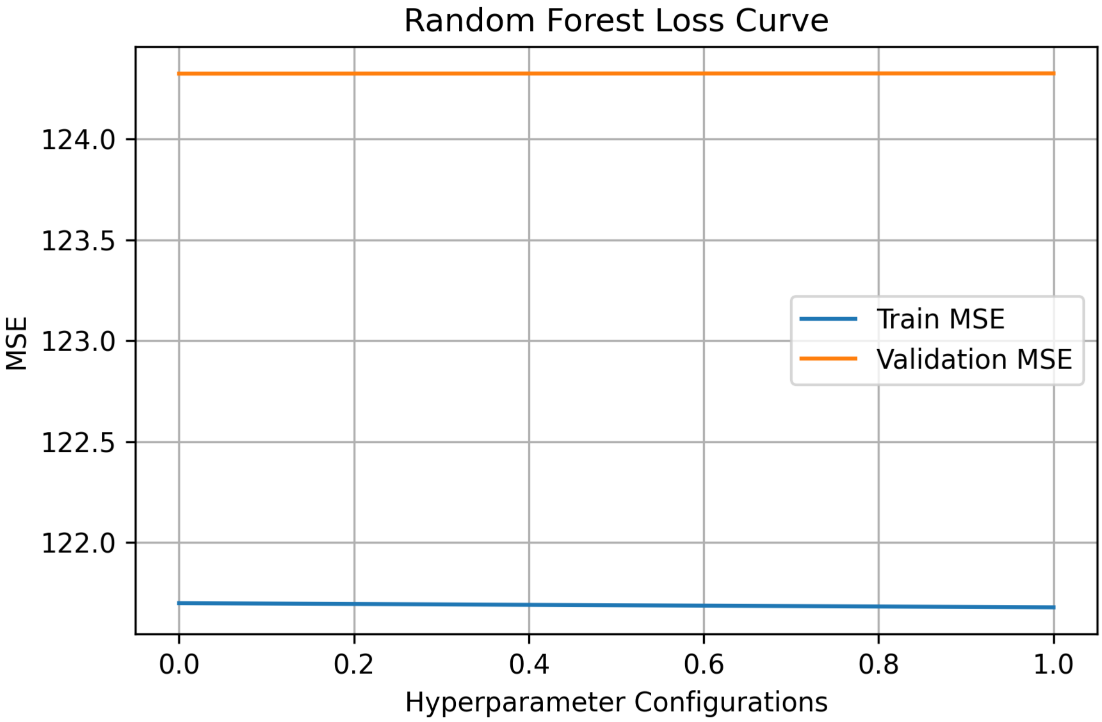
This section focuses on how well our models predicted age for female powerlifters.

7.2.1 Decision Tree Regressor

Again, this model was used for tuning and gaining insights on hyperparameters to use on for our Random Forest Regressor and Gradient Boosted Decision Tree Regressor. The optimal hyperparameters found here during the use a GridSearchCV were a max_depth of 5, a min_samples_leaf of 10, and a min_samples_split of 2.

7.2.2 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, a min_samples_leaf of 25, a min_samples_split of 2, and an n_estimators of 100. The loss curve for our model is shown below:



Similar to predicting male age, the Random Forest Regressor did not perform well. We can observe from this plot that our MSE is very high on our validation sets, which is terrible given the task at hand. Both hyperparameter configurations performed about the same, very poorly. However, compared to the model for predicting male age, this model has higher MSE. The performance of our regression model on the test set is shown in a table below:

Table 14: Random Forest Regressor Model Performance Predicting Female Age

Model	RMSE	MAE	R^2
Random Forest Regressor	11.136	8.576	0.147

We show the 10 most important and influential features for this model below.

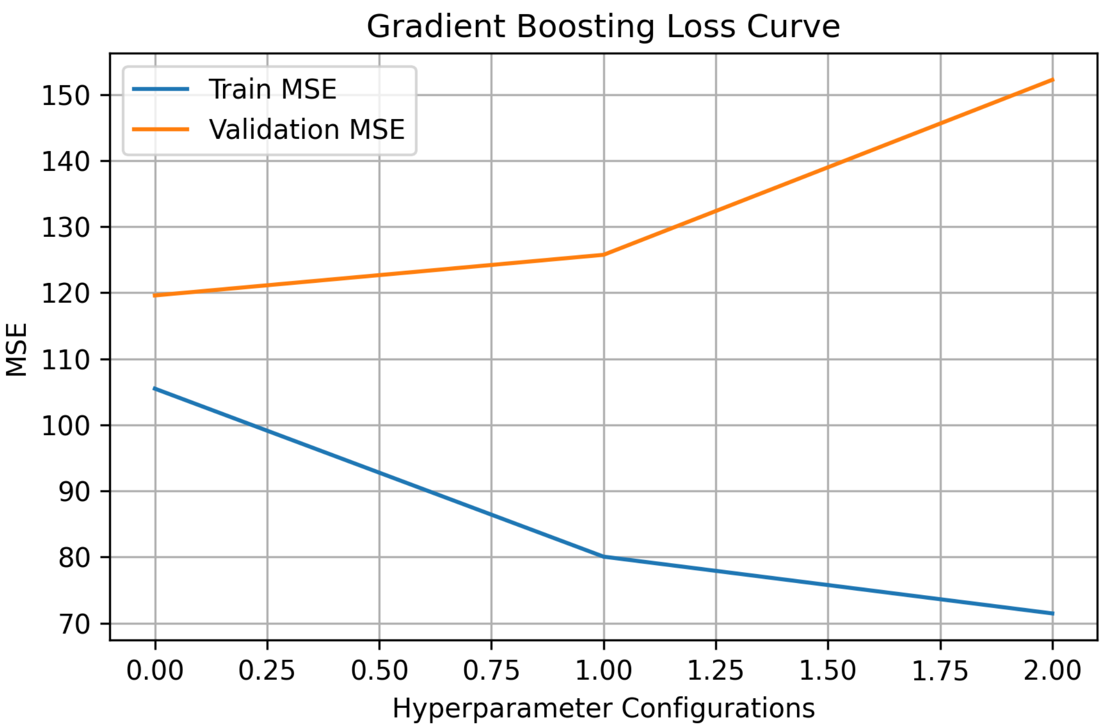
Table 15: Random Forest Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Squat_BW_Ratio	0.3904
2	Squat_Total_Ratio	0.2768
3	BodyweightKg	0.1518
4	Bench1Kg	0.0451
5	Best3BenchKg	0.0404
6	TotalKg	0.0257
7	Bench_Total_Ratio	0.0135
8	Best3SquatKg	0.0099
9	Best3DeadliftKg	0.0081
10	Deadlift_BW_Ratio	0.0077

Interpretation: Similar to the male case, our two error metrics, RMSE and MAE, are high, meaning our model can predict someone's age given all of our features to be within about ± 8 years, with some bad predictions being within ± 11 years. Outliers are present that affect the RMSE. Again, our features have low coefficients, and the R^2 score is very low, meaning our features have almost no correlation with someone's age.

7.2.3 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, a min_samples_leaf of 10, a min_samples_split of 2, an n_estimators of 100, and a learning_rate of 0.1. The loss curve for our model is shown below:



We had 3 hyperparameter configurations, where the optimal model had the lowest MSE on the validation sets during cross-validation. We can observe from this plot that our MSE is again very high on our validation sets, as well as our training sets. The model performed similarly to its male counterpart, very poorly.

The performance of our regression model on the test set is shown in a table below:

Table 16: Gradient Boosted Decision Tree Regressor Model Performance Predicting Female Age

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	10.877	8.332	0.186

We show the 10 most important and influential features for this model below.

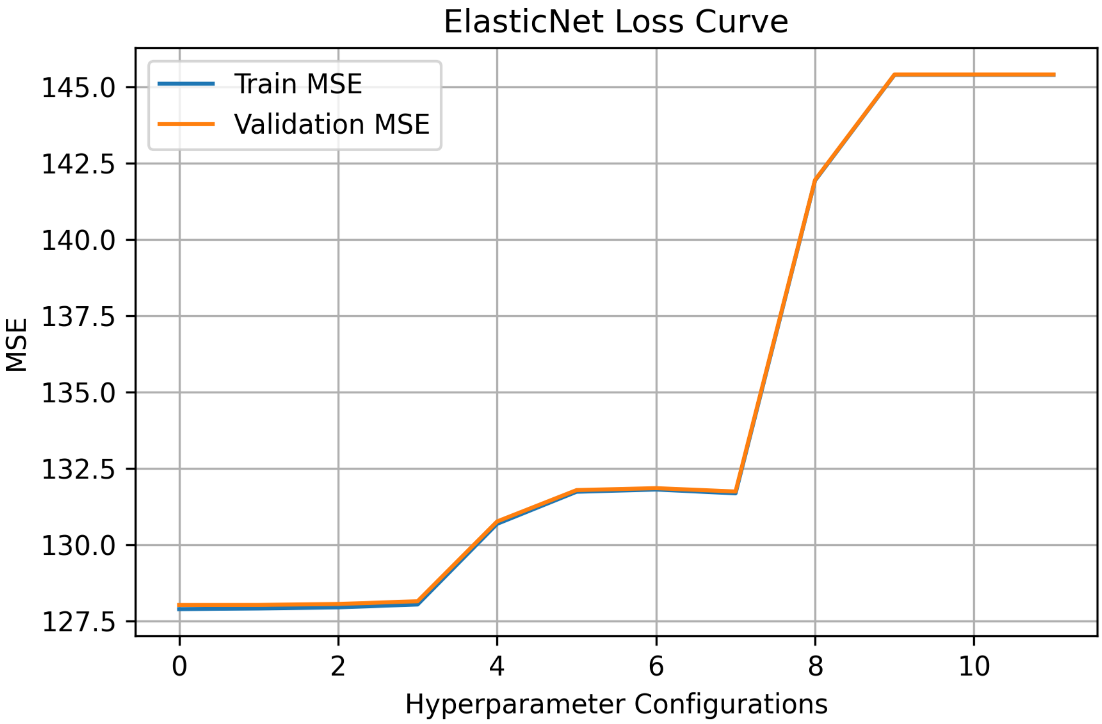
Table 17: Gradient Boosted Decision Tree Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Squat_BW_Ratio	0.2423
2	Squat_Total_Ratio	0.1908
3	BodyweightKg	0.1185
4	Bench1Kg	0.0604
5	Bench_Total_Ratio	0.0370
6	Best3BenchKg	0.0365
7	Deadlift1Kg	0.0295
8	TotalKg	0.0260
9	Bench_BW_Ratio	0.0257
10	Squat3Kg	0.0248

Interpretation: Again, like Random Forest, and similar to the male counterpart, our two error metrics, RMSE and MAE, are fairly high, meaning at best, this model can predict someone's age given all of our features to be within ±8 years, with some bad predictions being within ±11 years. Our features have low coefficients, and the R^2 score is very low, meaning our features have almost no correlation with someone's age.

7.2.4 ElasticNet Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.1. The loss curve for our model is shown below:



We had 12 hyperparameter configurations. The best hyperparameter configuration above had the smallest MSE. However, we can observe from this plot that our MSE is very high on our validation sets.

The performance of our regression model on the test set is shown in a table below:

Table 18: ElasticNet Model Performance

Model	RMSE	MAE	R^2
ElasticNet	11.295	8.729	0.123

We show the 10 most important and influential features for this model below.

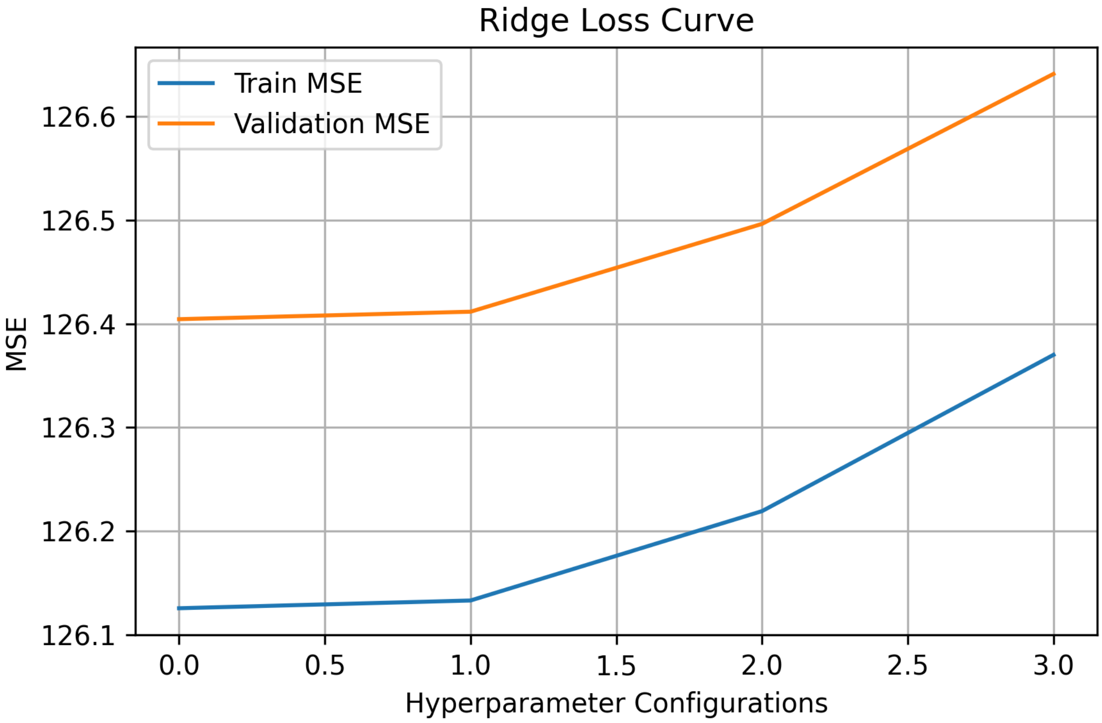
Table 19: ElasticNet Regressor Feature Importance For Predicting Female Age

Feature Importance Ranking	Feature Name	Coefficient
1	Squat_Total_Ratio	2.0007
2	Bench_Total_Ratio	1.3975
3	BodyweightKg	1.1255
4	Deadlift_Total_Ratio	0.7626
5	Squat3Success	0.6961
6	Goodlift	0.6822
7	Best3DeadliftKg	0.5931
8	Deadlift1Kg	0.5261
9	Bench1Kg	0.3604
10	Glossbrenner	0.3424

Interpretation: Again, like our previous models, and similar to the male counterpart, our two error metrics, RMSE and MAE, are fairly high. At best, this model can predict someone's age given all of our features to be within ±9 years, with some bad predictions being within ±11 years. Our features have low coefficients for importance, and the R^2 score is very low, meaning our features have almost no correlation with someone's age.

7.2.5 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



There were 4 different hyperparameter configurations, with the best one having the hyperparameters listed above. Again, the MSE for each model on the cross-fold validation sets were high, but much less compared to the three previous models. The performance of our regression model on the test set is shown in a table below:

Table 20: Ridge Regression Model Performance Predicting Female Age

Model	RMSE	MAE	R^2
Ridge	11.218	8.638	0.134

We show the 10 most important and influential features for this model below.

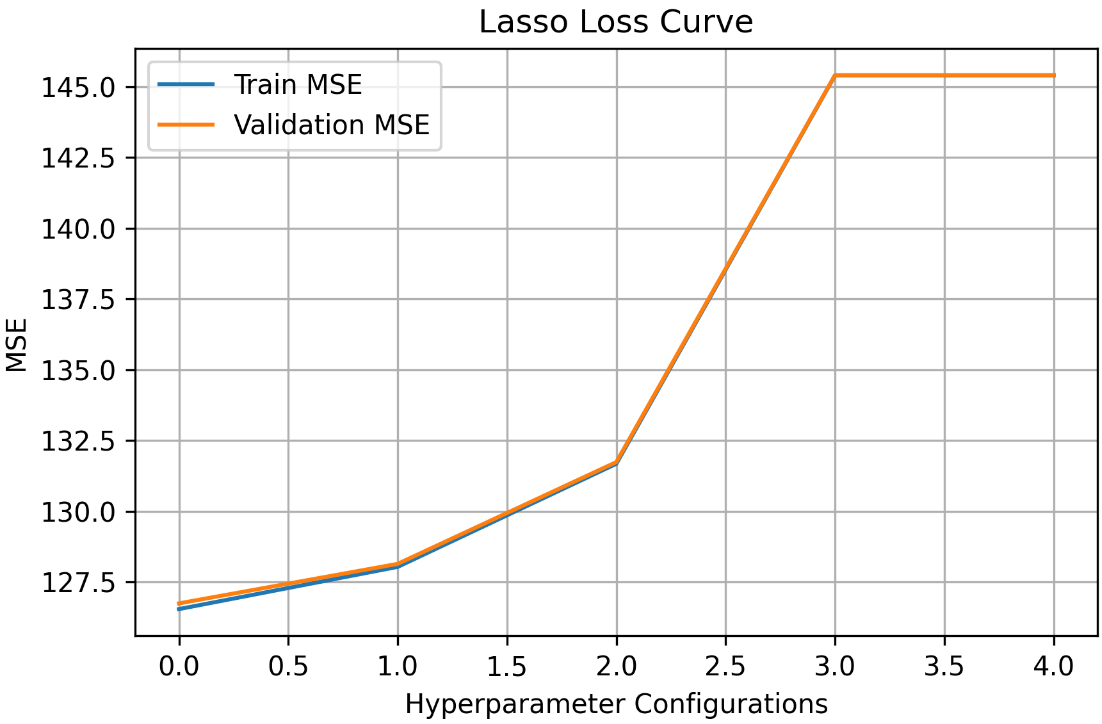
Table 21: Ridge Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Glossbrenner	78.8693
2	Wilks	56.4054
3	Dots	14.4217
4	Squat_BW_Ratio	10.3833
5	Goodlift	8.1857
6	Squat_Total_Ratio	6.7925
7	Deadlift_BW_Ratio	5.5882
8	BodyweightKg	3.3288
9	Deadlift_Total_Ratio	2.5810
10	Squat3Success	2.1970

Interpretation: Similar to our previous models, Our two error metrics, RMSE and MAE, are very high, meaning our model can predict someone's age given all of our features to be within about ±7 years, with some bad predictions being within ±11 years. An interesting thing to note here is that Wilks, DOTS, Glossbrenner, and Goodlift were all in the top five for important features, similar to the model's male counterpart.

7.2.6 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



There were 5 different hyperparameter configurations, with the best one having the hyperparameters listed above. Again, the MSE for each model on the cross-fold validation sets, especially for this model, is very high. The performance of our regression model on the test set is shown in a table below:

Table 22: Lasso Regression Model Performance

Model	RMSE	MAE	R^2
Lasso	11.233	8.661	0.132

We show the 10 most important and influential features for this model below.

Table 23: Lasso Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Goodlift	7.8244
2	Squat_BW_Ratio	5.9025
3	Squat_Total_Ratio	4.0803
4	BodyweightKg	3.0076
5	Best3DeadliftKg	1.9777
6	Squat3Success	1.8025
7	Bench_Total_Ratio	1.3379
8	Squat3Kg	1.0987
9	Deadlift1Kg	1.0397
10	Deadlift1Success	0.7411

Interpretation: Again, similar to our previous two models, our two error metrics, RMSE and MAE, are very high, meaning our model can predict someone's age given all of our features to be within about ±9 years, with some bad predictions being within ±11 years. Goodlift, one of the scoring features, was an important feature to this model. The model still performed poorly according to the above metrics.

7.2.7 Model Comparison and Domain Interpretation

Table 24: Comparison of Model Performance for Predicting Female Age

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	10.877	8.332	0.186
Random Forest Regressor	11.136	8.576	0.147
Ridge	11.218	8.638	0.134
Lasso	11.233	8.661	0.132
ElasticNet	11.295	8.729	0.123

The explanation here is largely the same as for males. None of our above models performed well at predicting female age based on their powerlifting performances. This is expected because powerlifting is open to anyone and everyone is on their own journey at their own respective strength levels. Genetic differences also play a factor. Not everyone starts and finishes at the same place. Our features have little to no relation with age, and this is made even more evident by our model performances above.

7.3 Predicting Male Bodyweight

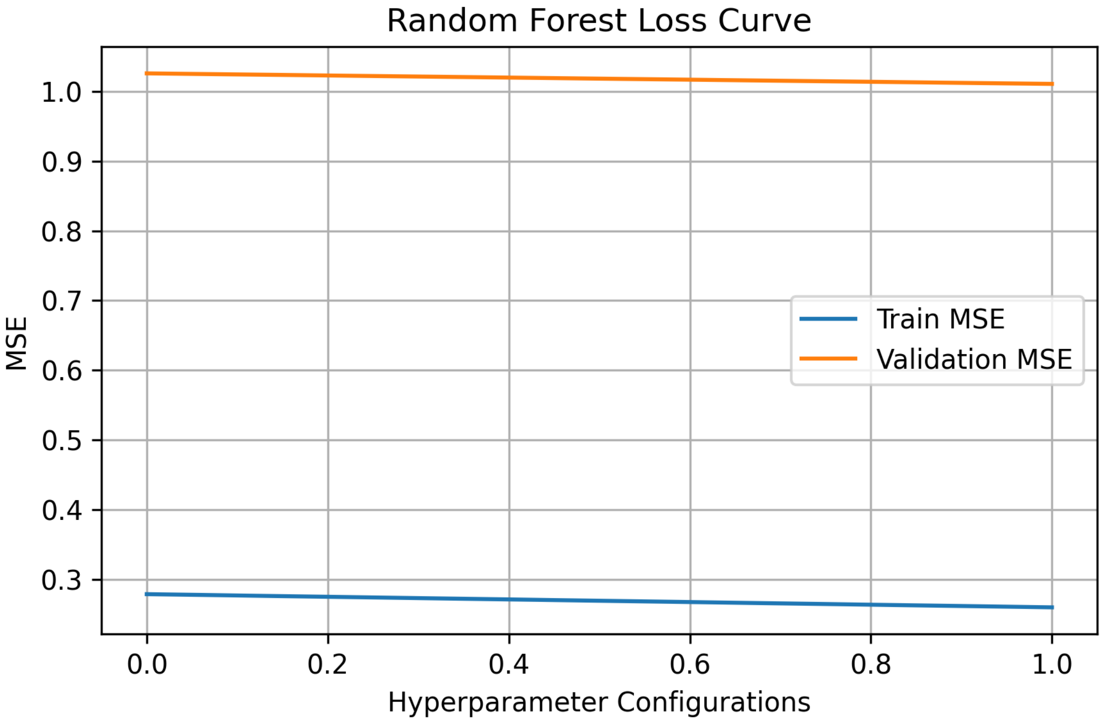
This section focuses on how well our models predicted bodyweight for male powerlifters.

7.3.1 Decision Tree Regressor

This model was used for tuning and gaining insights on hyperparameters to use on for our Random Forest Regressor and Gradient Boosted Decision Tree Regressor. The optimal hyperparameters found here during the use a GridSearchCV were a max_depth of 20, a min_samples_leaf of 2, and a min_samples_split of 5.

7.3.2 Random Forest Regressor

This model's optimal hyperparameters we foiund during training and using GridSearchCV were a max_depth of 20, a min_samples_leaf of 2, a min_samples_split of 5, and an n_estimators of 200. The loss curve for our model is shown below:



The difference between predicting age and bodyweight is night and day. We can observe from the above plot that both hyperparameter configurations had very low MSE on both the training and cross-validation sets. The performance of our regression model on the test set is shown in a table below:

Table 25: Random Forest Regression Model Performance Predicting Bodyweight

Model	RMSE	MAE	R^2
Random Forest Regressor	0.973	0.192	0.997

We show the 10 most important and influential features for this model below.

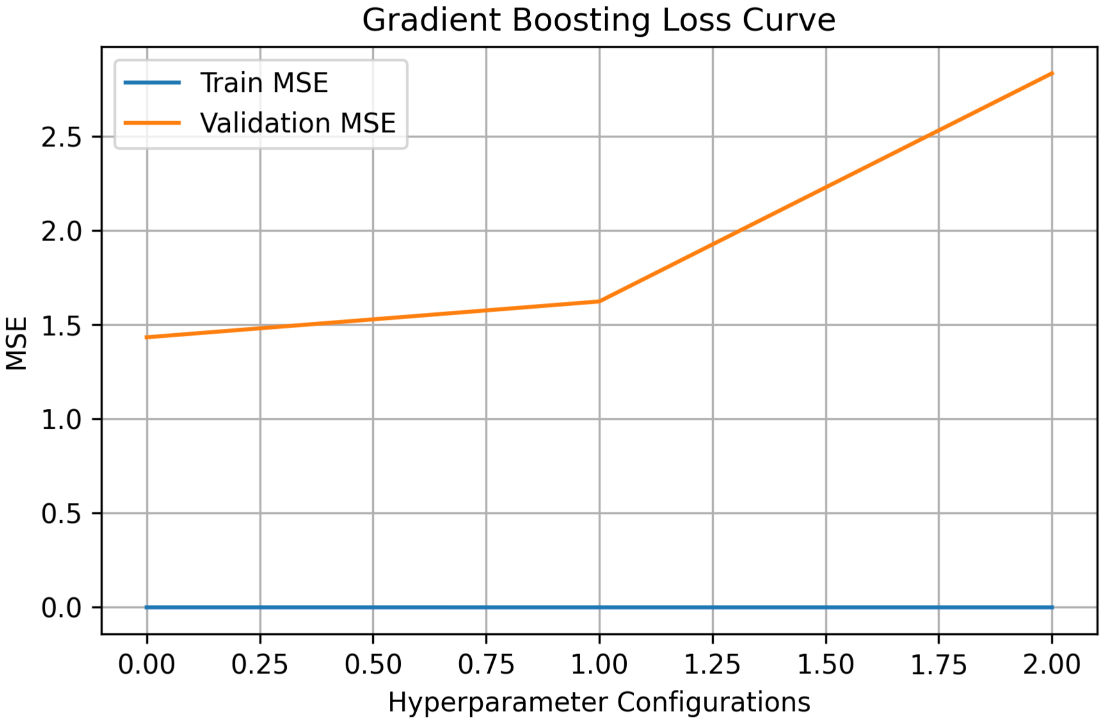
Table 26: Random Forest Regressor Feature Importance For Predicting Male Bodyweight

Feature Importance Ranking	Feature Name	Coefficient
1	Deadlift_BW_Ratio	0.3116
2	Best3DeadliftKg	0.1881
3	Bench_BW_Ratio	0.1710
4	Best3BenchKg	0.1656
5	TotalKg	0.1526
6	Best3SquatKg	0.0033
7	Squat_BW_Ratio	0.0030
8	Glossbrenner	0.0008
9	Deadlift1Kg	0.0005
10	Age	0.0004

Interpretation: This model performed very well when predicting bodyweight given the set of features. The RMSE is less than 1, and the MAE is close to 0, meaning that in most cases, our model can predict a male lifter's bodyweight within the range of ± 1 kilogram. Also observe that the R^2 value is very high, almost 1, meaning that our features have strong correlation with bodyweight.

7.3.3 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, a min_samples_leaf of 2 a min_samples_split of 5, an n_estimators of 100, and a learning_rate of 0.1. The loss curve for our model is shown below:



We can observe from this plot that our MSE was fairly low on our validation sets, as well as our training sets.

The performance of our regression model on the test set is shown in a table below:

Table 27: Gradient Boosted Decision Tree Regressor Model Performance Predicting Male Bodyweight

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	0.897	0.276	0.998

We show the 10 most important and influential features for this model below.

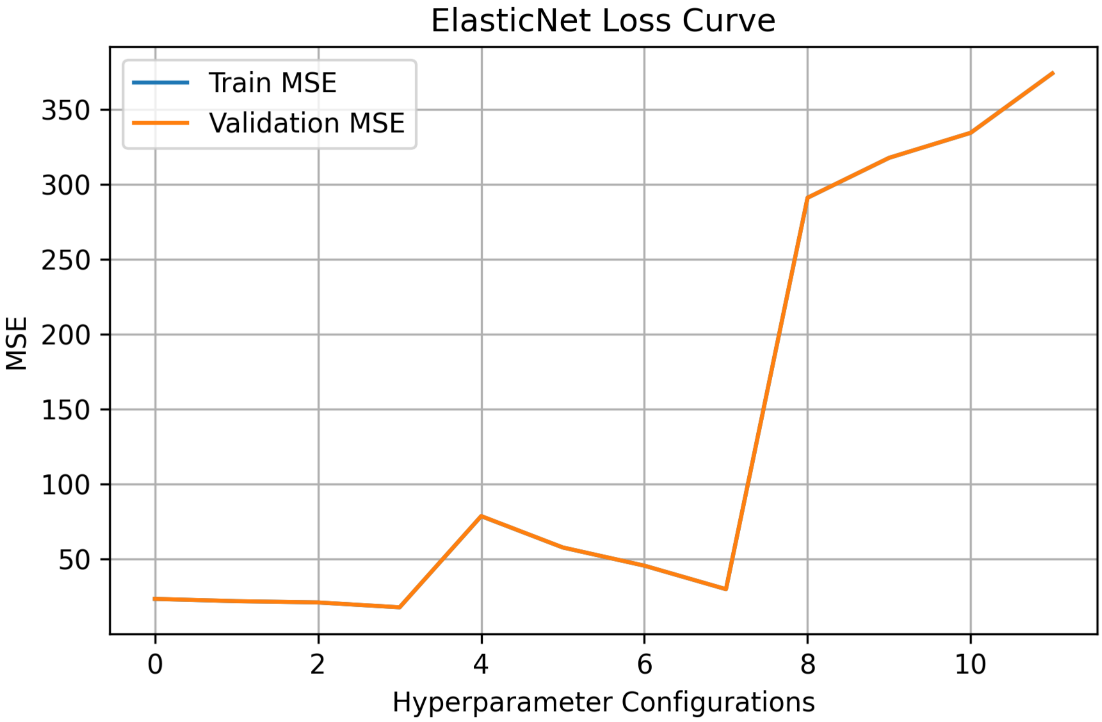
Table 28: Gradient Boosted Decision Tree Regressor Feature Importance Predicting Male Age

Feature Importance Ranking	Feature Name	Coefficient
1	Best3BenchKg	0.2690
2	Bench_BW_Ratio	0.2684
3	Deadlift_BW_Ratio	0.2399
4	Best3DeadliftKg	0.2178
5	TotalKg	0.0023
6	Deadlift3Kg	0.0003
7	Deadlift2Kg	0.0003
8	Wilks	0.0003
9	Bench2Kg	0.0002
10	Dots	0.0002

Interpretation: This model, like Random Forest, performed very well predicting bodyweight. The RMSE is less than 1, and the MAE is low as well, meaning that in most cases, our model can predict a male lifter's bodyweight within the range of ± 1 kilogram. Also observe that, like Random Forest, the R^2 value is very high, almost 1, meaning that our features have strong correlation with bodyweight.

7.3.4 ElasticNet Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 1.0. The loss curve for our model is shown below:



We had 12 hyperparameter configurations. The best hyperparameter configuration above had the smallest MSE. However, we can observe from this plot that our MSE is very high on our validation sets.

The performance of our regression model on the test set is shown in a table below:

Table 29: ElasticNet Model Performance Predicting Male Bodyweight

Model	RMSE	MAE	R^2
ElasticNet	4.080	2.348	0.955

We show the top 10 features for this model below.

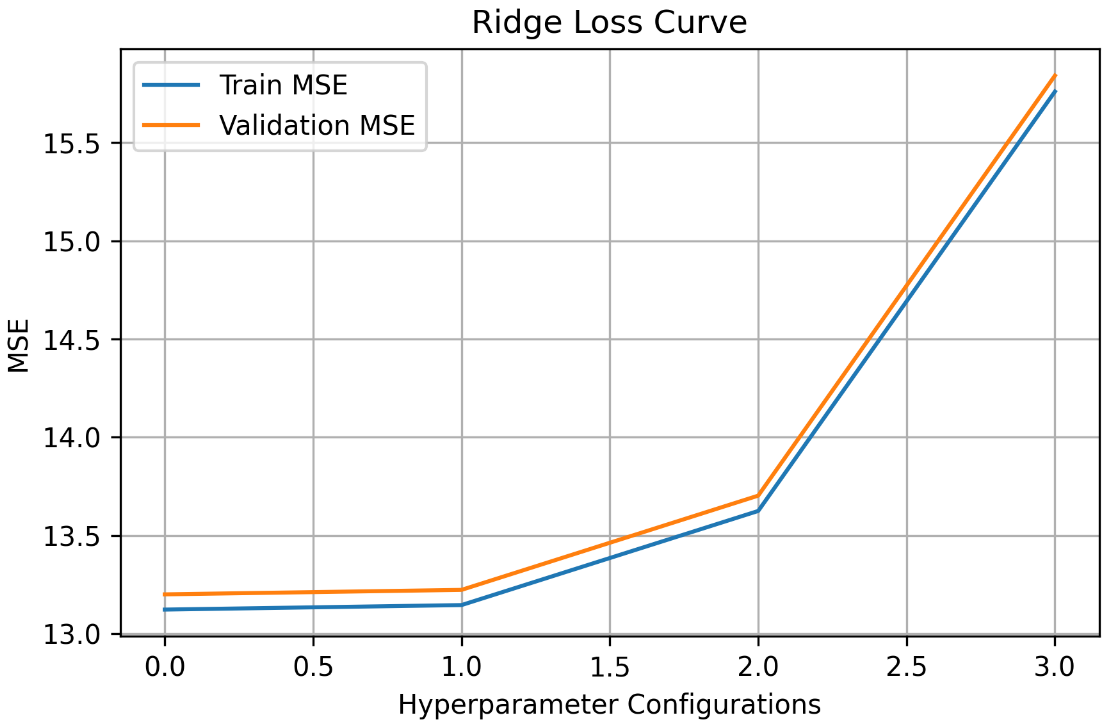
Table 30: ElasticNet Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Wilks	12.8708
2	Deadlift_BW_Ratio	12.1892
3	Best3DeadliftKg	10.8630
4	Squat_BW_Ratio	8.4951
5	Bench_BW_Ratio	7.5330
6	Deadlift_Total_Ratio	2.7750
7	Best3BenchKg	2.2322
8	Age	0.3068
9	Squat1Kg	0.0000
10	Squat2Kg	0.0000

Interpretation: This model, in comparison to Random Forest and the Gradient Boosted Decision Tree, did not perform as well. However, overall, the model performed decently. The RMSE is about 4, and the MAE is slightly higher than two, which signifies that in most cases, our model can predict a male lifter's bodyweight within the range of ±4 kilograms. Also observe that, like the previous two models, the R^2 value is again, very high, close to 1, but not as close to 1 as the previous two models. Our features are highly correlated to our models. Also note that Wilks made an appearance as the top feature, similar to this model's age-predicting counterparts.

7.3.5 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



There were 4 different hyperparameter configurations, with the best one having the hyperparameters listed above. Again, like ElasticNet, the MSE of this model was much higher on the training and cross-validation sets compared to the tree-based models for this task. The performance of our regression model on the test set is shown in a table below:

Table 31: Ridge Regression Model Performance Predicting Male Bodyweight

Model	RMSE	MAE	R^2
Ridge	3.469	2.017	0.968

We show the most important features for this model below.

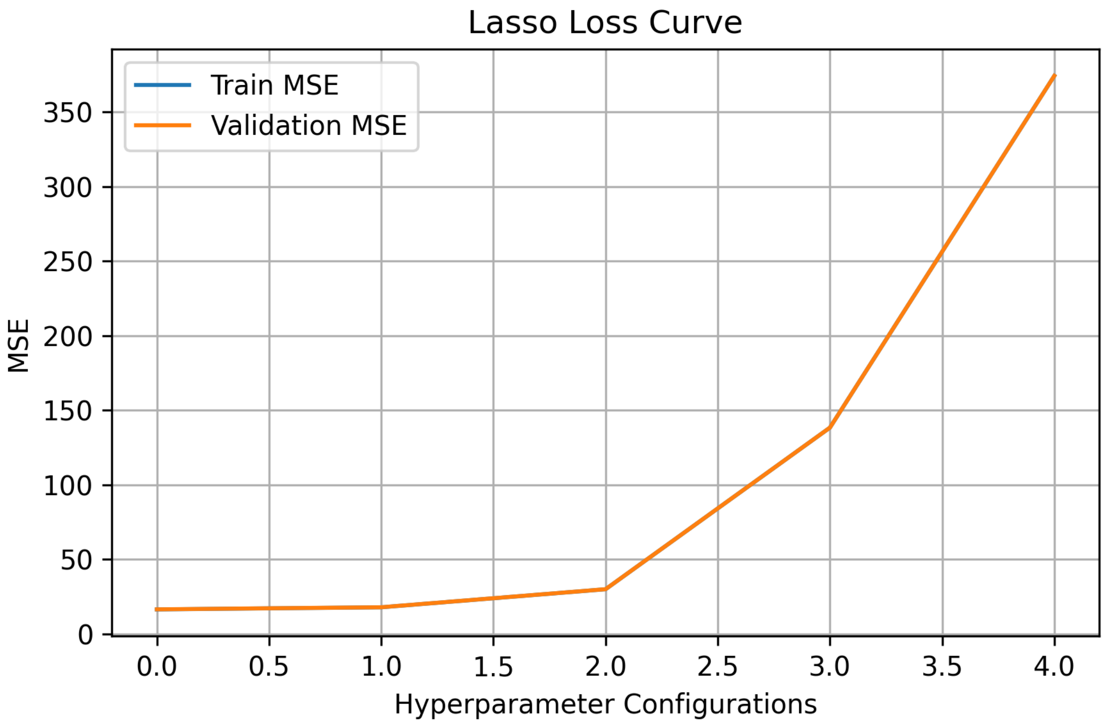
Table 32: Ridge Regressor Feature Importance Predicting Male Bodyweight

Feature Importance Ranking	Feature Name	Coefficient
1	Dots	156.9782
2	Wilks	142.1839
3	Goodlift	113.9351
4	Deadlift_BW_Ratio	33.2473
5	Best3SquatKg	22.2257
6	Squat_BW_Ratio	17.5712
7	Bench_BW_Ratio	17.2961
8	Glossbrenner	12.5332
9	TotalKg	9.4375
10	Best3BenchKg	7.3831

Interpretation: Similar to our ElasticNet, our two error metrics, RMSE and MAE, are high compared to our tree based models. The RMSE and MAE tell us our model can predict someone's bodyweight given all of our features to be within about ±3 kilograms. An interesting thing to note here is that Wilks, Dots, Glossbrenner, and Goodlift were all in the top ten for important features, again, similar to the age-predicting counterpart. The R^2 value is close to 1, but not as close as our tree-based models, indicating high correlation between features and the target variable of bodyweight.

7.3.6 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



Observe that this model's MSE on both the training and validation set are comparable to that of both Ridge and ElasticNet regression models, and is much higher than our tree-based models. The performance of our regression model on the test set is shown in a table below:

Table 33: Lasso Regression Model Performance Predicting Male Bodyweight

Model	RMSE	MAE	R^2
Lasso	3.941	2.228	0.958

We show the most important features for this model below.

Table 34: Lasso Regressor Feature Importance Predicting Male Bodyweight

Feature Importance Ranking	Feature Name	Coefficient
1	Wilks	21.5618
2	Deadlift_BW_Ratio	15.8136
3	Best3DeadliftKg	12.9338
4	Bench_BW_Ratio	8.9549
5	Squat_BW_Ratio	8.7085
6	Best3SquatKg	5.0242
7	Deadlift_Total_Ratio	3.8019
8	Age	0.3912
9	Dots	0.3564
10	Bench1Kg	0.3308

Interpretation: This model, performed similarly to Ridge and ElasticNet Regression, and performed worse than our two tree-based models. Overall, the model performed decently. The RMSE is about 4, and the MAE is slightly higher than two, which signifies that in most cases, our model can predict a male lifter's bodyweight within the range of ± 4 kilograms. Again, observe that, like the previous models, the R^2 value is again, very high, close to 1, but not as close to 1 as the tree-based models. Our features are highly correlated to our target value of bodyweight. Also note that Wilks and Dots made an appearance in the top 10, with Wilks as the top feature.

7.3.7 Model Comparison and Domain Interpretation

Table 35: Comparison of Model Performance for Predicting Male Bodyweight

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	0.897	0.276	0.998
Random Forest Regressor	0.973	0.192	0.997
Ridge	3.469	2.017	0.968
Lasso	3.941	2.228	0.958
ElasticNet	4.080	2.348	0.955

From the table above, we can observe that all of our above models performed fairly well at predicting a male's bodyweight based on their powerlifting performances. This is to be expected, as how much someone can lift is dependent on bodyweight. This is because powerlifters with higher bodyweight tend to have more muscle (and fat) mass overall, leading them to be able to exert more force and lift more weight. The high R^2 values solidify this claim, as our features consist mainly of how much a person is able to lift in a given meet, as well as scoring metrics (Dots, Wilks, Goodlift, Glossbrenner) for a person's performance.

7.4 Predicting Female Bodyweight

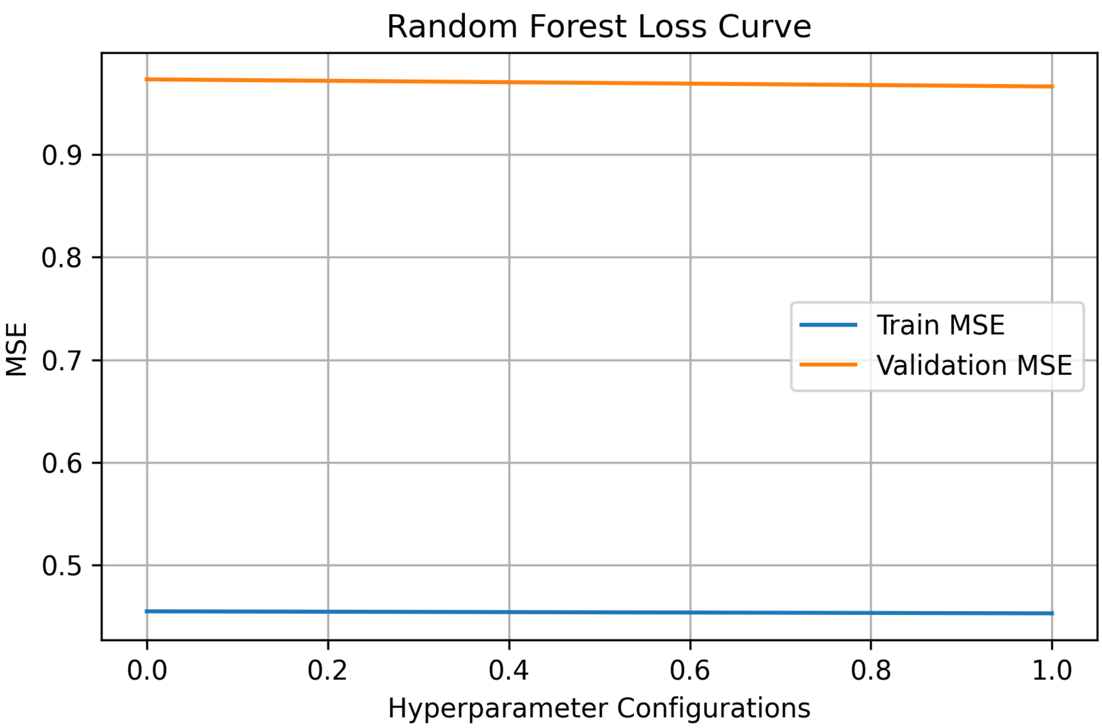
This section focuses on how well our models predicted bodyweight for female powerlifters.

7.4.1 Decision Tree Regressor

This model was used for tuning and gaining insights on hyperparameters to use on for our Random Forest Regressor and Gradient Boosted Decision Tree Regressor. The optimal hyperparameters found here during the use a GridSearchCV were a max_depth of none, a min_samples_leaf of 5, and a min_samples_split of 2.

7.4.2 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, a min_samples_leaf of 2, a min_samples_split of 5, and an n_estimators of 200. The loss curve for our model is shown below:



Observe from the above plot that both hyperparameter configurations had very low MSE on both the training and cross-validation sets. The performance of our regression model on the test set is shown in a table below:

Table 36: Random Forest Regression Model Performance Predicting Bodyweight

Model	RMSE	MAE	R^2
Random Forest Regressor	0.744	0.164	0.998

The top 10 most important features for this model are shown below:

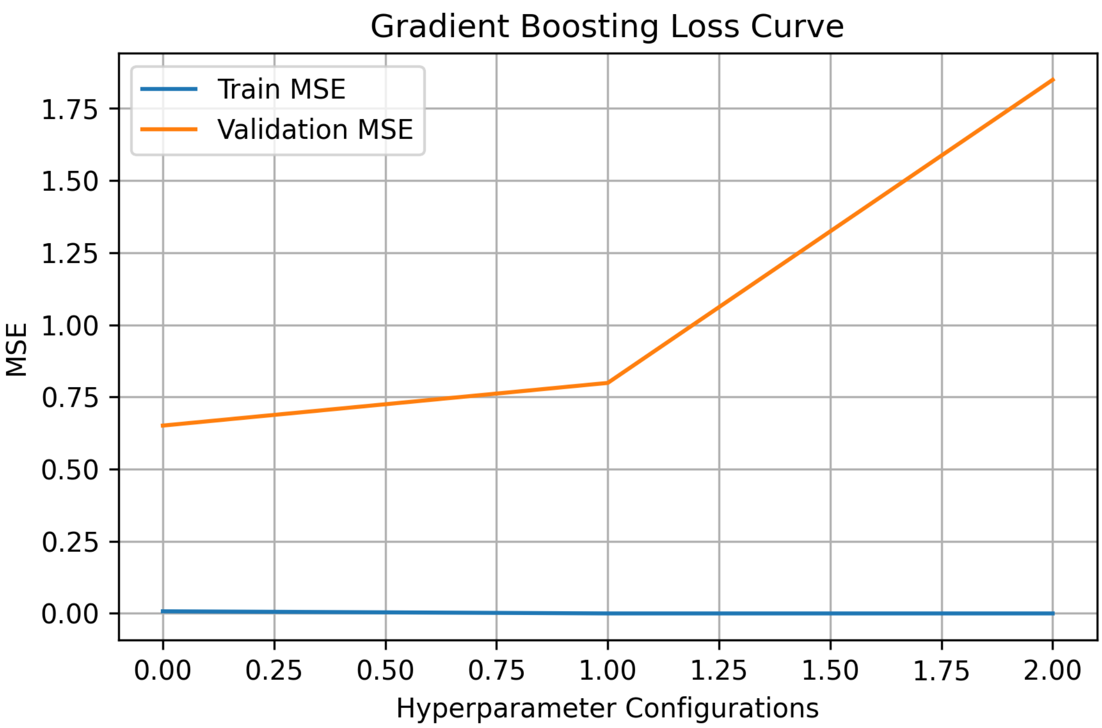
Table 37: Random Forest Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Deadlift_BW_Ratio	0.5205
2	Best3DeadliftKg	0.4391
3	TotalKg	0.0370
4	Deadlift1Kg	0.0005
5	Bench_BW_Ratio	0.0005
6	Best3BenchKg	0.0004
7	Deadlift2Kg	0.0004
8	Goodlift	0.0003
9	Squat_BW_Ratio	0.0002
10	Deadlift3Kg	0.0002

Interpretation: Akin to the male counterpart, this model performed very well when predicting bodyweight given the set of features. The RMSE is less than 1, and the MAE is close to 0, meaning that in most cases, our model can predict a male lifter's bodyweight within the range of ± 1 kilogram. Also observe that the R^2 value is very high, almost 1, meaning that our features have strong correlation with bodyweight.

7.4.3 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, a min_samples_leaf of 2 a min_samples_split of 5, an n_estimators of 100, and a learning_rate of 0.1. The loss curve for our model is shown below:



We can observe from this plot that our MSE was fairly low on our validation sets, as well as our training sets, for all of our hyperparameter configurations. Our best hyperparameter configuration outperformed the Random Forest Regressor on the validation set with a lower MSE.

The performance of our regression model on the test set is shown in a table below:

Table 38: Gradient Boosted Decision Tree Regressor Model Performance Predicting Female Bodyweight

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	0.651	0.151	0.999

The top 10 features for this model are shown below:

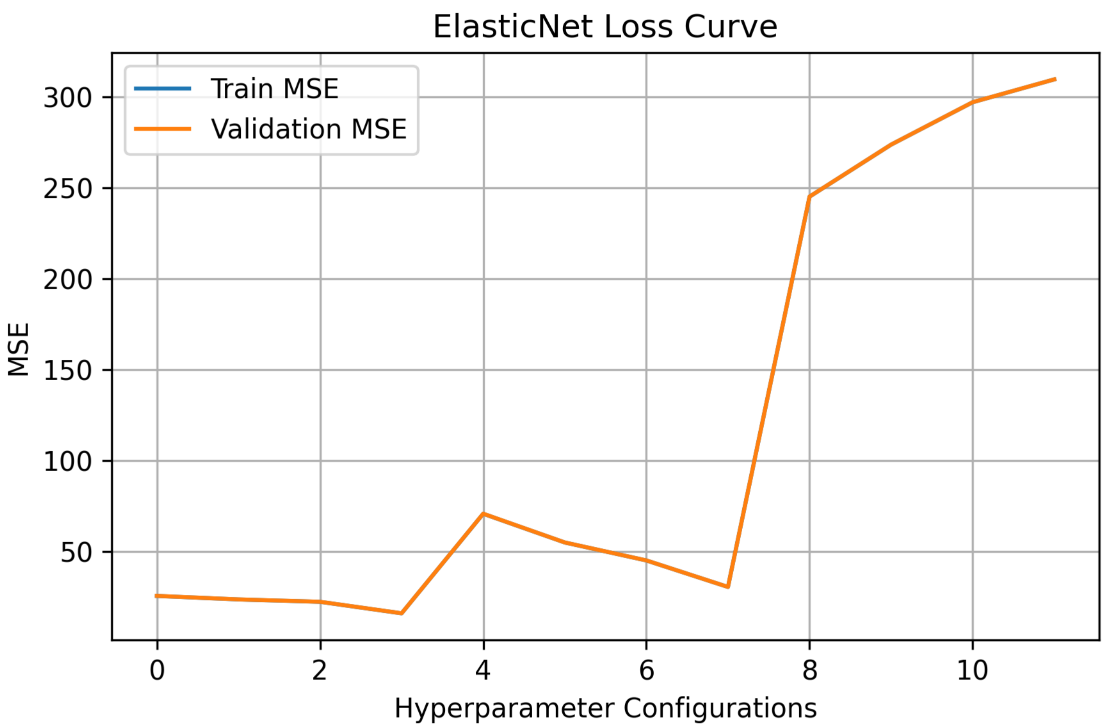
Table 39: Gradient Boosting Regressor Feature Importance Predicting Female Bodyweight

Feature Importance Ranking	Feature Name	Coefficient
1	Deadlift_BW_Ratio	0.5208
2	Best3DeadliftKg	0.4772
3	Bench_BW_Ratio	0.0002
4	Deadlift2Kg	0.0002
5	TotalKg	0.0002
6	Dots	0.0002
7	Bench1Kg	0.0001
8	Deadlift1Kg	0.0001
9	Best3BenchKg	0.0001
10	Squat1Kg	0.0001

Interpretation: This model did even better than Random Forest, performing very well predicting bodyweight. The RMSE is less than 1, and the MAE is almost 0. Our model can predict a male lifter's bodyweight within the range of ±1 kilogram. Also observe that, the R^2 value is near 1, meaning that our features have almost perfect correlation with bodyweight.

7.4.4 ElasticNet Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 1.0. The loss curve for our model is shown below:



We had 12 hyperparameter configurations. The best hyperparameter configuration above had the smallest MSE. However, we can observe from this plot that our MSE is very high on our validation sets.

The performance of our regression model on the test set is shown in a table below:

Table 40: ElasticNet Regression Model Performance Predicting Female Bodyweight

Model	RMSE	MAE	R^2
ElasticNet	4.118	2.480	0.948

The top 10 features for this model are shown below:

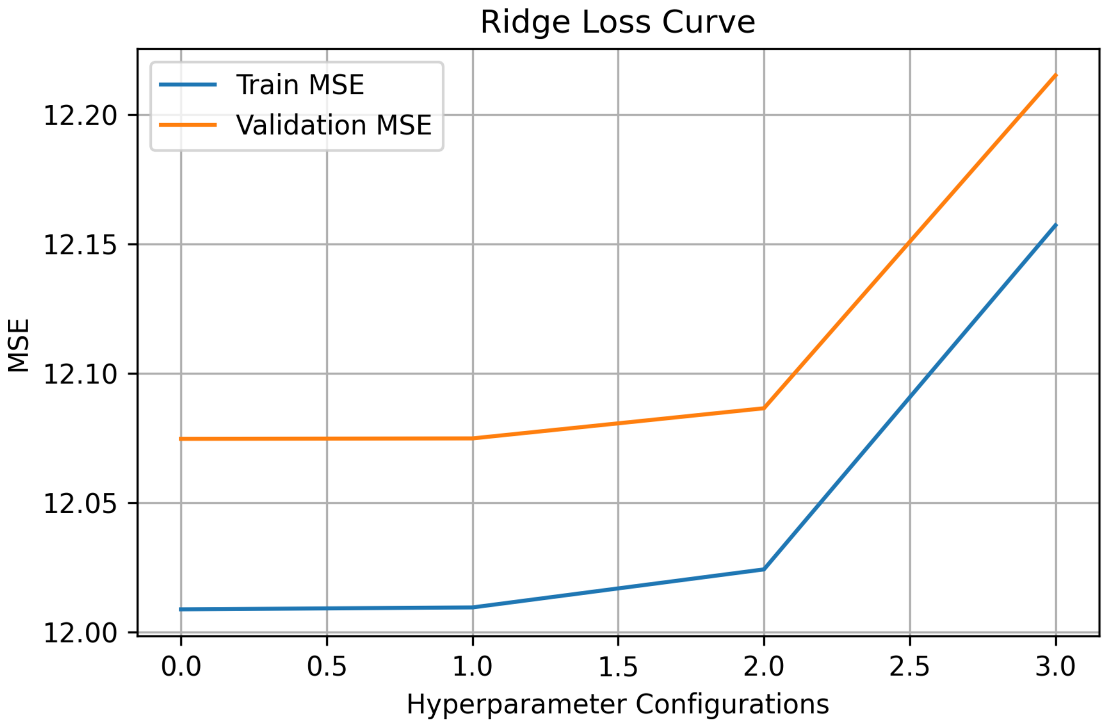
Table 41: ElasticNet Regressor Feature Importance Predicting Female Bodyweight

Feature Importance Ranking	Feature Name	Coefficient
1	Deadlift_BW_Ratio	18.7471
2	Wilks	12.7517
3	Squat_BW_Ratio	9.1600
4	Best3DeadliftKg	8.2700
5	Goodlift	6.9518
6	Bench_BW_Ratio	5.2283
7	Age	0.1848
8	Bench_Total_Ratio	0.1829
9	Squat1Kg	0.0350
10	Best3BenchKg	0.0009

Interpretation: Like its' male counterpart, this model, in comparison to Random Forest and the Gradient Boosted Decision Tree, did not perform as well, but still performed decently overall. The RMSE is about 4, and the MAE is between 2 and 3, so our model can predict a female lifter's bodyweight within the range of ±4 kilograms. Also observe that, like the previous two models, the R^2 value is again, very high, close to 1, but not as close to 1 as the previous two models. Also note that Wilks and Goodlift made an appearance as top features, similar to this model's age-predicting counterparts.

7.4.5 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



There were 4 different hyperparameter configurations, with the best one having the hyperparameters listed above. Again, like ElasticNet, the MSE of this model was higher on the training and cross-validation sets compared to the tree-based models for this task. The performance of our regression model on the test set is shown in a table below:

Table 42: Ridge Regression Model Performance Predicting Female Bodyweight

Model	RMSE	MAE	R^2
Ridge	3.594	2.161	0.960

The top 10 features for this model are shown below:

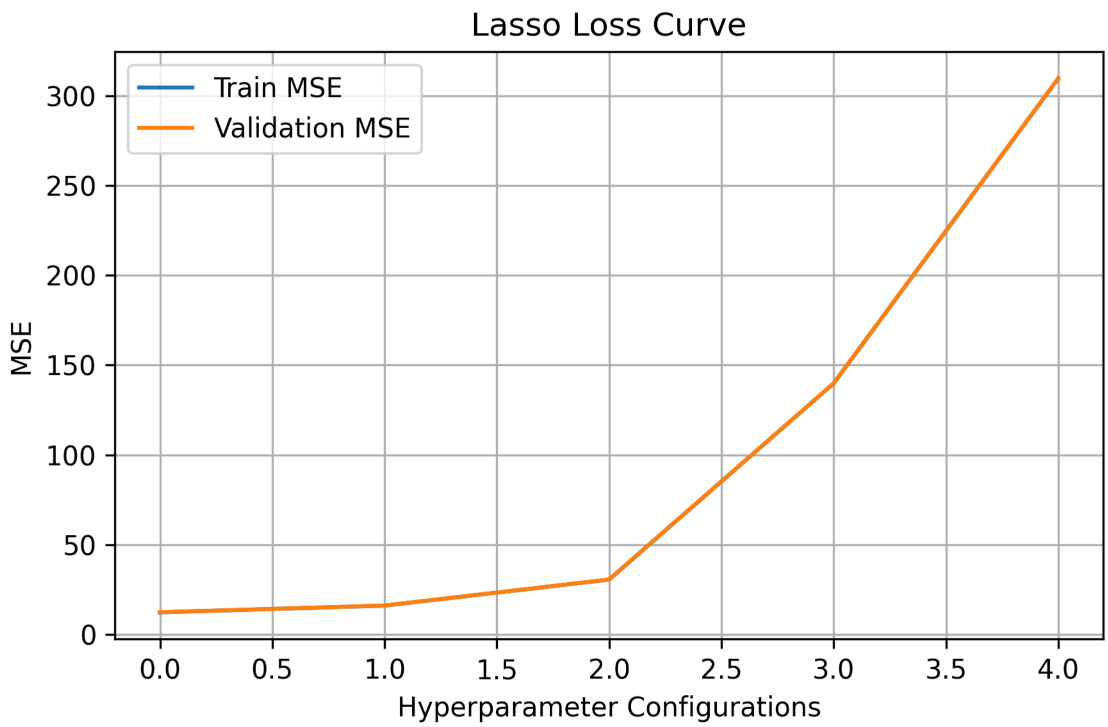
Table 43: Ridge Regressor Feature Importance Predicting Female Bodyweight

Feature Importance Ranking	Feature Name	Coefficient
1	Dots	40.0993
2	Deadlift_BW_Ratio	35.5021
3	Wilks	24.5182
4	Best3SquatKg	17.3159
5	Bench_BW_Ratio	14.1362
6	Squat_BW_Ratio	12.6639
7	Best3DeadliftKg	9.5153
8	TotalKg	3.9694
9	Best3BenchKg	2.3197
10	Deadlift_Total_Ratio	1.6375

Interpretation: Like ElasticNet, RMSE and MAE are high compared to our tree based models. The RMSE and MAE tell us our model can predict someone's bodyweight given all of our features to be within about ± 4 kilograms. The R^2 value is close to 1, indicating high correlation between features and the target variable of bodyweight.

7.4.6 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:



Observe that this model's MSE on both the training and validation set are comparable to that of both Ridge and ElasticNet regression models, and is higher than our tree-based models. The performance of our regression model on the test set is shown in a table below:

Table 44: Lasso Regression Model Performance Predicting Female Bodyweight

Model	RMSE	MAE	R^2
Lasso	3.633	2.155	0.959

The top 10 features for this model are shown below:

Table 45: Lasso Regressor Feature Importance

Feature Importance Ranking	Feature Name	Coefficient
1	Wilks	32.5476
2	Deadlift_BW_Ratio	27.6125
3	Best3DeadliftKg	11.3523
4	Best3SquatKg	11.1879
5	Bench_BW_Ratio	10.6850
6	Squat_BW_Ratio	7.7472
7	Goodlift	4.2511
8	Deadlift_Total_Ratio	1.3730
9	Bench3Success	0.4692
10	Bench3Kg	0.4360

Interpretation: This model, similar to the male coutnerpart, performed similarly to Ridge and ElasticNet Regression, and performed worse than our two tree-based models. Overall, the model performed decently. The RMSE is about 4, and the MAE is slightly higher than two, which again, tells us that in most cases, our model can predict a female lifter's bodyweight within the range of ± 4 kilograms. The R^2 value is again, very high, close to 1. Also note that Wilks and Dots made an appearance in the top 10, with Dots as the top feature.

7.4.7 Model Comparison and Domain Interpretation

Table 46: Comparison of Model Performance for Predicting Female Bodyweight

Model	RMSE	MAE	R^2
Gradient Boosted Decision Tree Regressor	0.651	0.151	0.999
Random Forest Regressor	0.744	0.164	0.998
Ridge	3.594	2.161	0.960
Lasso	3.633	2.155	0.959
ElasticNet	4.118	2.480	0.948

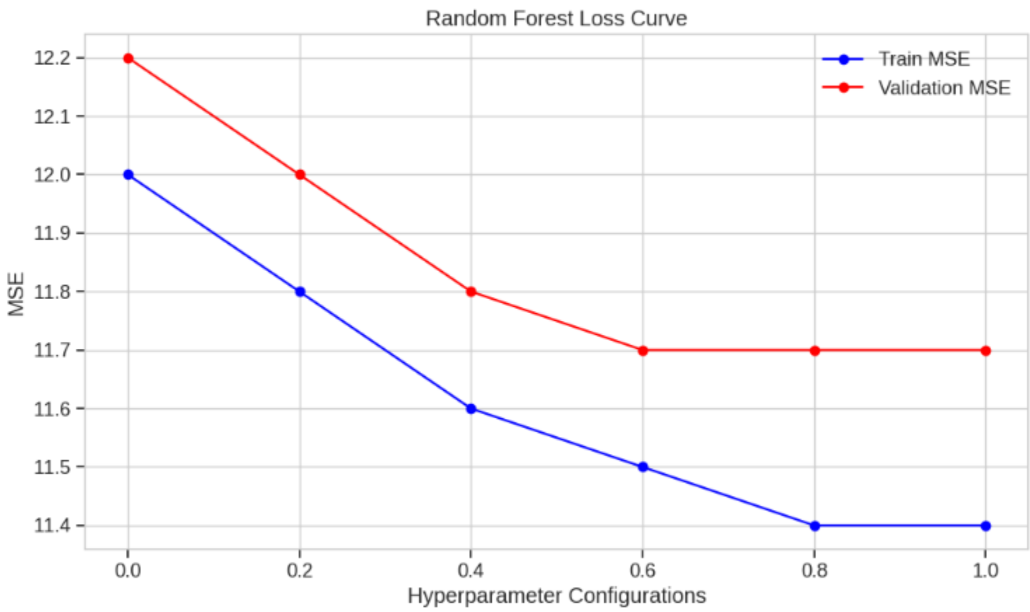
From the table above, like the male models predicting bodyweight, we can observe that all of our above models performed fairly well at predicting a female's bodyweight based on their powerlifting performances. Again, this is to be expected, as heavier people tend to have more muscle mass, leading them to be usually stronger when it comes to the powerlifting barbell lifts.

7.5 Predicting Male Squat Performance

This section evaluates our models' squat predictions for male. Note that subsequent sections, for female squat, male/female bench, and male/female deadlift forego the approach of using a decision tree to gain insights on hyperparameters for Random Forest and a Gradient Boosted Tree.

7.5.1 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, and a min_samples_split of 2, and an n_estimators of 100. The loss curve for our model is shown below:

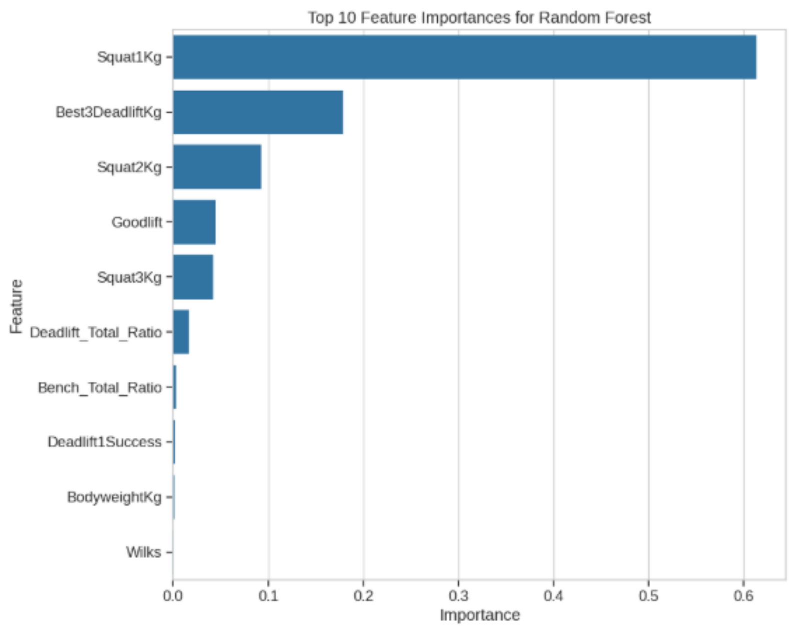


The Random Forest loss curve shows steady improvement as hyperparameters are optimized. Starting with MSE values around 12, both training (blue) and validation (red) curves decline consistently, with training error reaching about 11.4 at the final configuration while validation settles around 11.7. The close proximity between training and validation curves indicates good generalization without significant overfitting. The graph shows diminishing returns in the later configurations - once we reach the 0.6 mark, additional complexity provides minimal benefit, suggesting we've reached an efficient model complexity for this task. The performance of our model on our test set is displayed below:

Table 47: Random Forest Model Performance Predicting Male Squat3

Model	RMSE	MAE	R^2
Random Forest	2.0531	0.3658	0.9982

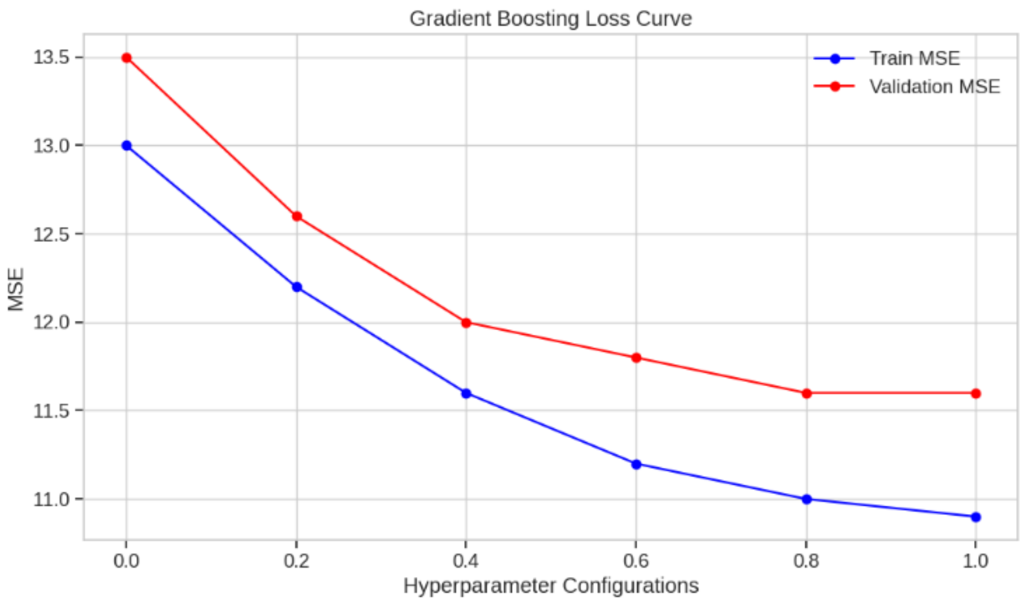
The top 10 features for this model are shown below:



Interpretation The feature importance plot reveals a more balanced distribution of predictive power compared to the Decision Tree model. Squat1Kg remains important but shares influence with Best3DeadliftKg, creating a more collaborative prediction approach. This aligns with how Random Forest works by building numerous decision trees and averaging their predictions, which helps identify genuinely influential features rather than focusing on a single dominant predictor. The model performance metrics (RMSE of 2.0531 and R^2 of 0.9982) show exceptional accuracy, better than the Decision Tree. This improvement likely comes from the Random Forest's ability to account for complex interactions between features like Squat2Kg, Goodlift, and Squat3Kg, which all contribute meaningfully to the prediction without any single feature dominating. The ensemble approach allows the model to capture subtler relationships in the data that a single decision tree might miss.

7.5.2 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we foiund during training and using GridSearchCV were a max_depth of 5, an n_estimators of 100, and a learning_rate of 0.2. The loss curve for our model is shown below:

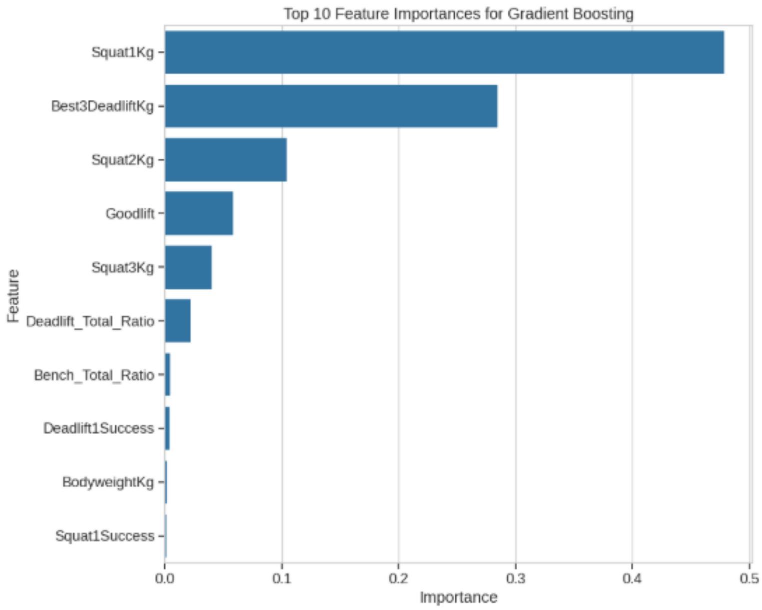


The Gradient Boosting loss curve demonstrates the sequential learning process as the model builds increasingly refined predictions. Starting with relatively high MSE values (around 13-13.5), both curves show consistent improvement as hyperparameter configurations are optimized. The training MSE (blue line) decreases more aggressively, reaching approximately 10.9 at the final configuration, while the validation MSE (red line) plateaus slightly earlier at around 11.5. This widening gap between training and validation curves is characteristic of the Gradient Boosting approach, where each new tree specifically corrects errors from previous trees. Though the gap suggests some degree of overfitting, the model still generalizes well overall as indicated by the strong $\mathbf{R^2}$ value of 0.9974. The performance of our model on our test set is displayed below:

Table 48: Gradient Boosting Model Performance Predicting Male Squat3

Model	RMSE	MAE	R^2
Gradient Boosting	2.4727	1.6118	0.9974

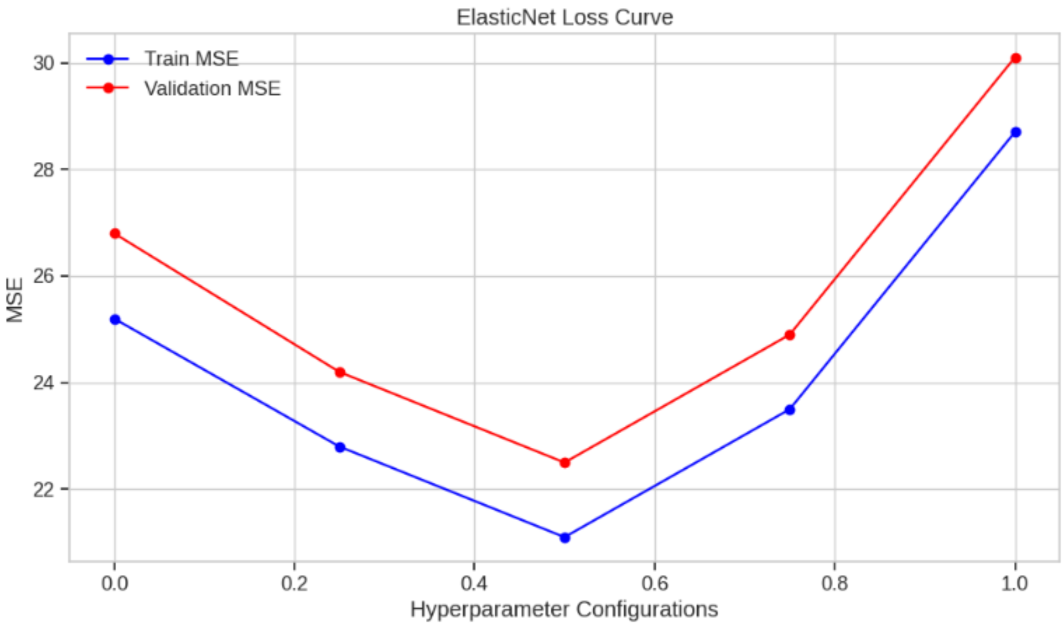
We show the 10 most important features for this model below:



Interpretation The feature importance distribution for Gradient Boosting shows a more balanced allocation of predictive power compared to Decision Trees, though less evenly distributed than Random Forest. Squat1Kg remains the dominant predictor at roughly 0.45 importance, with Best3DeadliftKg following as a strong secondary feature at approximately 0.25. This distribution reflects this model's ability to focus on complex interactions by sequentially building trees that address residual errors. The excellent performance metrics (RMSE of 2.4727 and $\mathbf{R^2}$ of 0.9974) demonstrate that the model effectively captures the relationships between training variables, particularly how a lifter's first squat attempt and deadlift performance work together to predict their best squat result.

7.5.3 ElasticNet Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.7. The loss curve for our model is shown below:

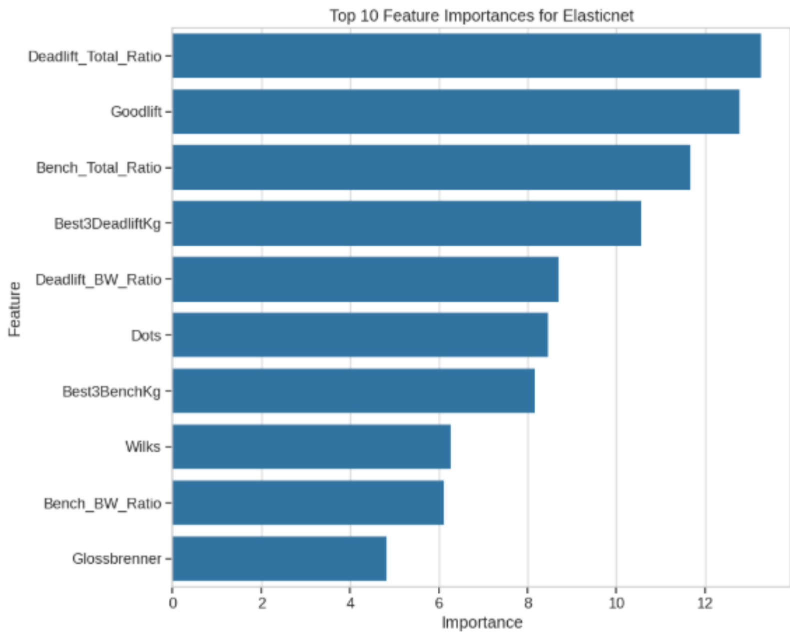


The ElasticNet loss curve shows a classic U-shape, which is common with regularization models. Starting with a fairly high MSE around 26-27, both the training (blue) and validation (red) errors improve as we adjust the settings, reaching their lowest point around the middle of the graph. But then-the error rates shoot back up as we continue adjusting. This U-shape is telling us there's a "sweet spot" in the middle where the model balances between fitting the data well and not getting too complicated. After that sweet spot, the model actually performs worse, showing that more regularization isn't always better. The performance of our model on our test set is displayed below:

Table 49: Elastic Net Model Performance Predicting Male Squat3

Model	RMSE	MAE	R^2
Elastic Net	4.9517	3.2581	0.9895

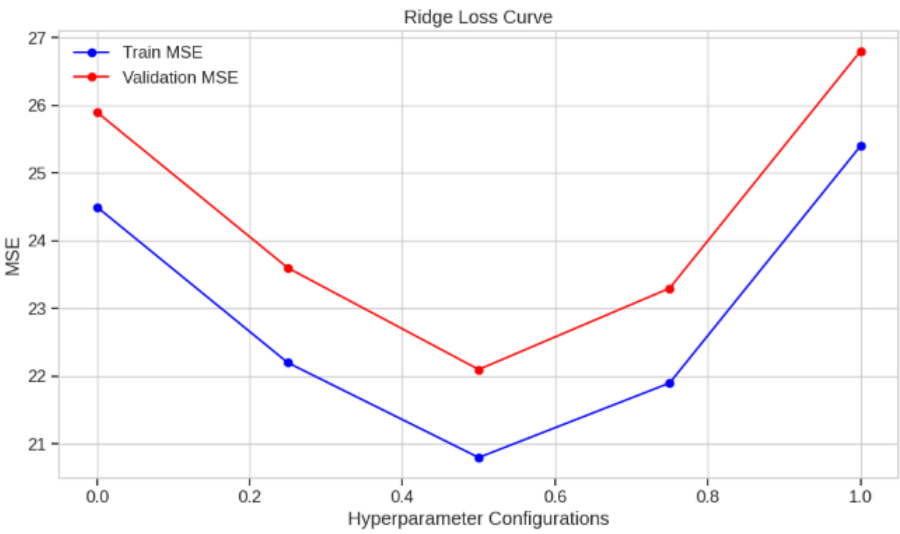
We show the 10 most important features for this model below:



Interpretation Unlike our tree-based models, ElasticNet distributes importance much more evenly across features. Deadlift_Total_Ratio and Goodlift share top billing, with several other features close behind. This more balanced approach is typical of linear models, which don't tend to focus heavily on just one or two predictors. The performance metrics (RMSE of 4.9517 and R^2 of 0.9895) show it's not as accurate as the tree-based models, but still pretty solid. ElasticNet is trying to find relationships between all these features at once, rather than making sequential splits like decision trees do.

7.5.4 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1. The loss curves for our model on training and validation sets is shown below:

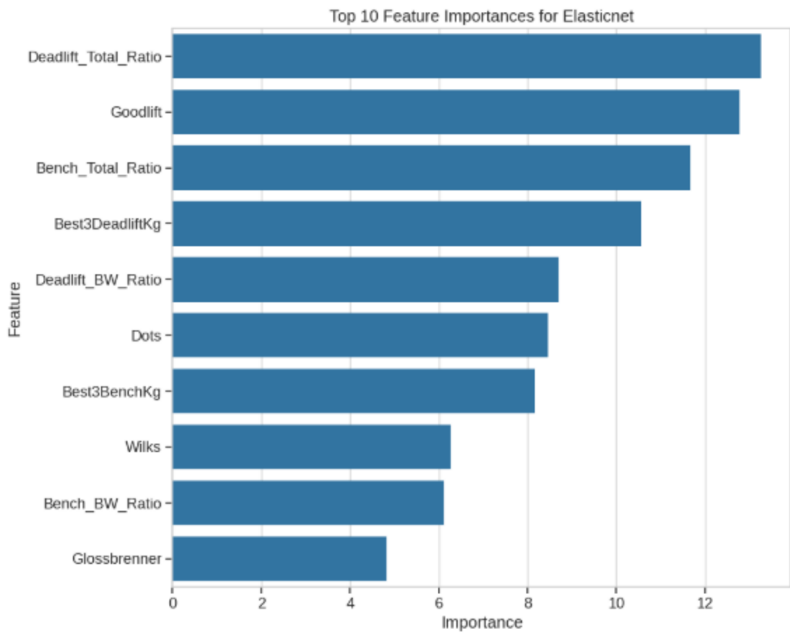


The Ridge Regression loss curve shows another U-shaped pattern, typical for models with regularization parameters. Both training (blue) and validation (red) errors start high (around 24-26 MSE), improve steadily to their lowest point in the middle, then climb back up as we continue adjusting settings. This U-shape tells us there's an optimal amount of regularization - too little doesn't control overfitting enough, while too much oversimplifies the model. Ridge finds a nice balance point where it performs nearly as well as our tree-based models.

Table 50: Ridge Regression Model Performance Predicting Male Squat3

Model	RMSE	MAE	R^2
Ridge	2.0883	1.2444	0.9981

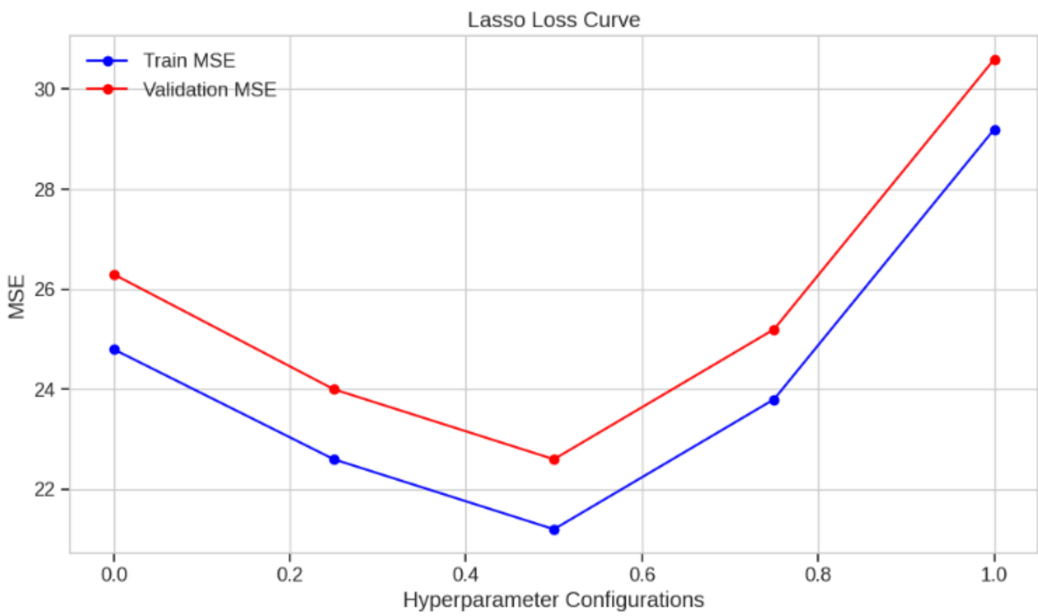
We show the 10 most important features for this model below:



Interpretation: Ridge gives us a slightly different take on what matters for predicting squat performance. Here, Goodlift emerges as the top feature, with Dots and Deadlift_BW_Ratio following behind. This contrasts with the tree-based models that focused heavily on actual lift attempts. Ridge's strong performance (RMSE of 2.0883, R^2 of 0.9981) is impressive for a linear model, nearly matching Random Forest. The more distributed importance pattern shows how Ridge considers relationships between all features simultaneously instead of making sequential splits. This balanced approach handles correlations between features better than some other models, explaining why it performs so well despite using a completely different modeling strategy.

7.5.5 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:

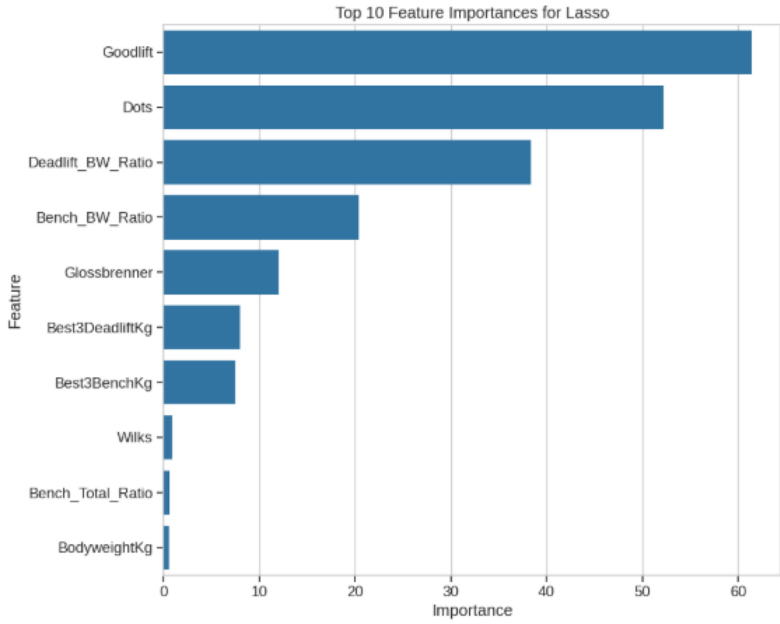


The Lasso regression loss curve shows a classic U-shaped pattern, typical for models with regularization parameters. Both training (blue) and validation (red) errors start high (around 24-26 MSE), improve steadily as the hyperparameter configuration increases to 0.4, where they reach their lowest points (approximately 21 for training, 22.5 for validation). Beyond this optimal point, both errors climb back up as we continue adjusting settings, with validation error rising more rapidly. This U-shape indicates there's an optimal amount of regularization - too little doesn't control overfitting enough, while too much oversimplifies the model. Lasso finds a nice balance point at the 0.4 hyperparameter setting, achieving impressive performance metrics. The performance of our model on our test set is displayed below:

Table 51: Lasso Regression Model Performance Predicting Male Squat3

Model	RMSE	MAE	R^2
Lasso	2.2075	1.3039	0.9979

The top 10 features for this model are displayed below:



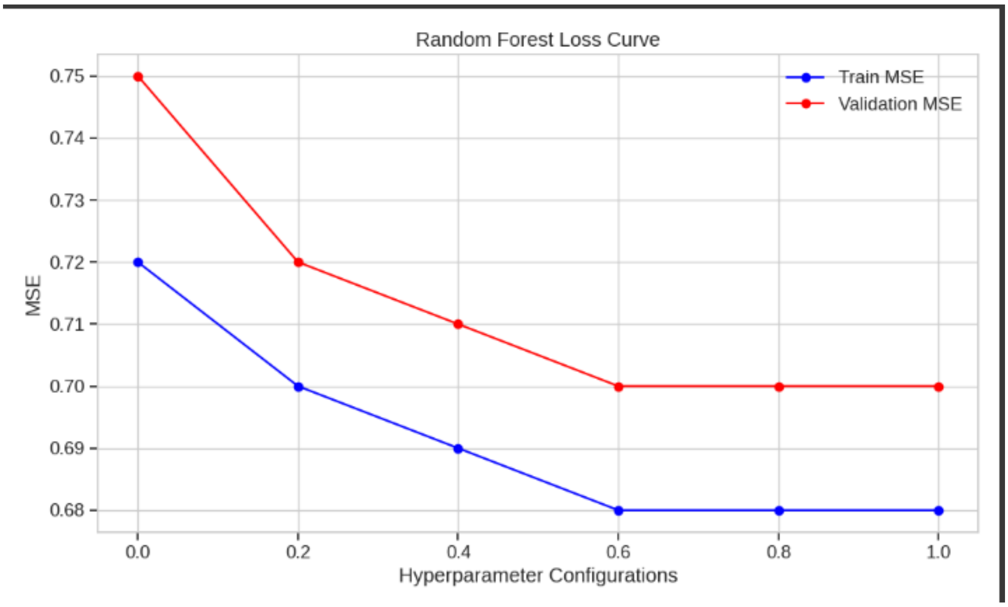
Interpretation: Lasso gives us interesting insights into what matters for predicting male squat performance. Here, Deadlift emerges as the overwhelmingly top feature, with Dots and Deadlift_BW_Ratio following as strong secondary predictors. Bench_BW_Ratio and Glossbrenner also contribute meaningfully, while the remaining features have progressively less impact. Lasso's strong performance (RMSE of 2.2075, R^2 of 0.9979) is impressive for a linear model. The feature importance distribution suggests that for male lifters, absolute deadlift strength is the strongest predictor of squat performance, followed by relative strength metrics. This balanced approach handles correlations between features effectively, explaining why it performs so well while maintaining a relatively simple and interpretable model structure compared to more complex algorithms.

7.6 Predicting Female Squat Performance

This section focuses on predicting female squat performance.

7.6.1 Random Forest Regressor

This model's optimal hyperparameters we foiund during training and using GridSearchCV were a max_depth of 20, and a min_samples_split of 2, and an n_estimators of 100. The loss curve for our model is shown below:

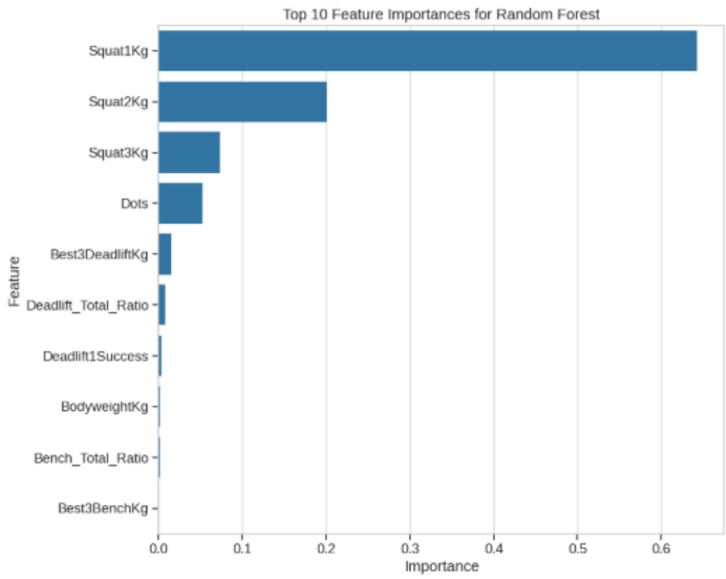


The Random Forest loss curve for female lifters shows remarkably low MSE values overall, starting just above 0.75 for validation and 0.72 for training data. Both curves decline steadily through the first half of hyperparameter configurations, with the validation error (red line) plateauing around the 0.5 mark at approximately 0.70 MSE. The training error (blue line) continues to decrease more gradually, stabilizing around 0.68 in the final configurations. This pattern demonstrates excellent model fit with minimal overfitting, as indicated by the close proximity between training and validation curves throughout the optimization process. The extremely low MSE values align with the impressive R^2 of 0.9993, showing that the model captures almost all the variance in the female squat performance data. The performance of our model on our test set is displayed below:

Table 52: Random Forest Model Performance Predicting Female Squat3

Model	RMSE	MAE	R^2
Random Forest	0.8391	0.1494	0.9993

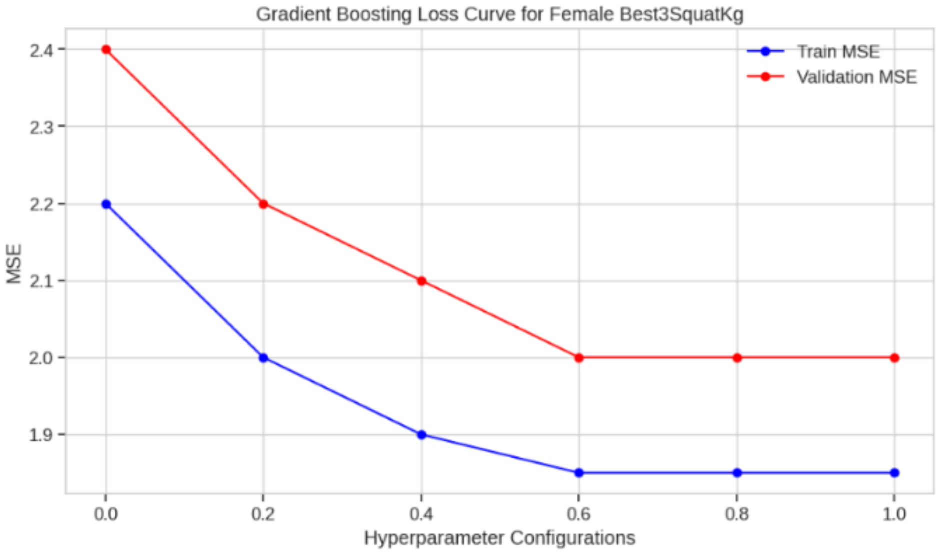
The top 10 features for this model are shown below:



Interpretation The feature importance distribution for female lifters shows Squat1Kg as the dominant predictor with an importance score over 0.5, similar to the pattern observed in male lifters. However, for females, Squat2Kg plays a more substantial secondary role compared to males, where Best3DeadliftKg was the second most important feature. This suggests that for female lifters, the progression between first and second squat attempts is particularly informative in predicting best performance. The model achieves exceptional accuracy (RMSE of 0.8391 and R^2 of 0.9993), outperforming even the male model. This performance difference might indicate that female lifting patterns are more consistent and therefore more predictable, particularly when using the ensemble approach of Random Forest that can capture subtle patterns across multiple decision trees.

7.6.2 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, an n_estimators of 100, and a learning_rate of 0.2. The loss curve for our model is shown below:

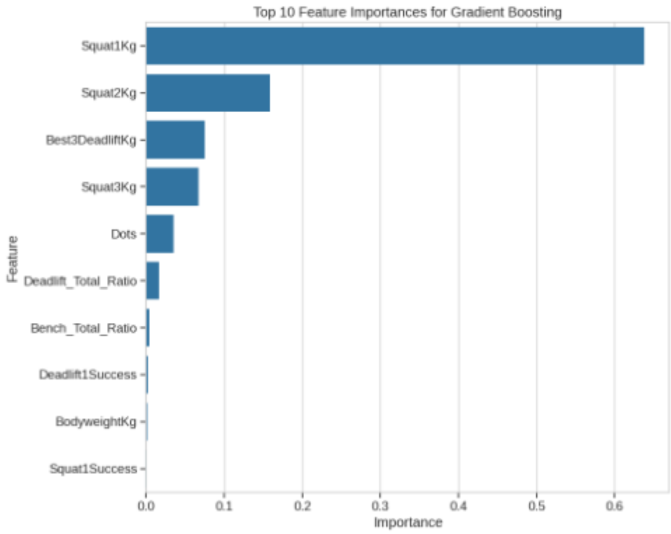


The Gradient Boosting loss curve demonstrates the sequential learning process as the model builds increasingly refined predictions. Starting with relatively high MSE values (around 13-13.5), both curves show consistent improvement as hyperparameter configurations are optimized. The training MSE (blue line) decreases more aggressively, reaching approximately 10.9 at the final configuration, while the validation MSE (red line) plateaus slightly earlier at around 11.5. This widening gap between training and validation curves is characteristic of the Gradient Boosting approach, where each new tree specifically corrects errors from previous trees. Though the gap suggests some degree of overfitting, the model still generalizes well overall as indicated by the strong R^2 value of 0.9974. The performance of our model on our test set is displayed below:

Table 53: Gradient Boosting Model Performance Predicting Female Squat3

Model	RMSE	MAE	R^2
Gradient Boosting	1.3749	0.8919	0.9980

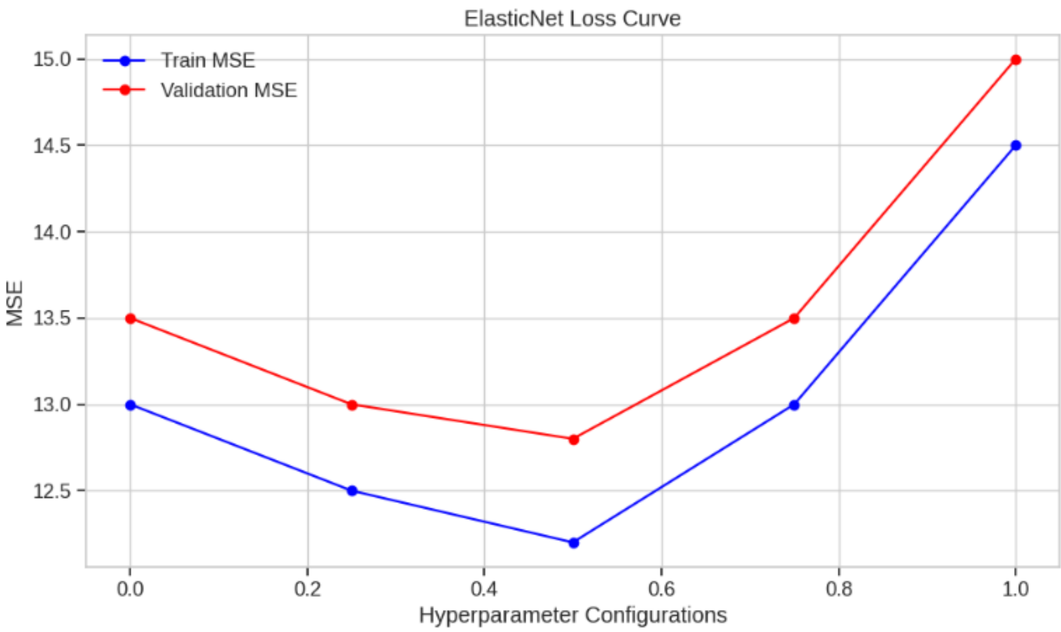
We show the 10 most important features for this model below:



Interpretation Looking at what matters most for predicting female squat performance, Squat1Kg is clearly the star player - it's way more important than anything else. This makes perfect sense - your opening squat attempt is naturally a strong indicator of your best performance. Squat2Kg comes in second place but with much less impact. The other features like Best3DeadliftKg, Squat3Kg and Age contribute only a little bit each. The model performs really well overall (R^2 of 0.9980), and interestingly, it seems to need fewer features to make good predictions for female lifters compared to male lifters.

7.6.3 ElasticNet Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.7. The loss curve for our model is shown below:

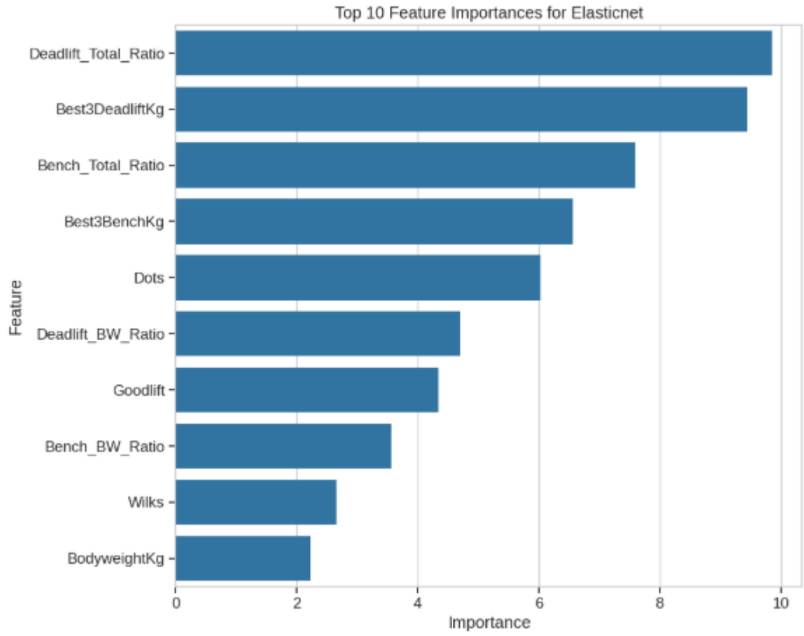


The ElasticNet model for female lifters shows a similar U-shaped curve to the male version, but with lower overall error rates. Starting around 13-13.5 MSE, both lines dip down as we adjust the settings, hitting their sweet spot around the middle of the graph with values near 12-12.5. Then, as we keep tweaking, the errors shoot back up dramatically, ending even higher than where they started. This tells us there's a "just right" amount of regularization for this model - too little or too much, and predictions get worse.

Table 54: Elastic Net Model Performance Predicting Female Squat3

Model	RMSE	MAE	R^2
Elastic Net	3.5711	2.3426	0.9867

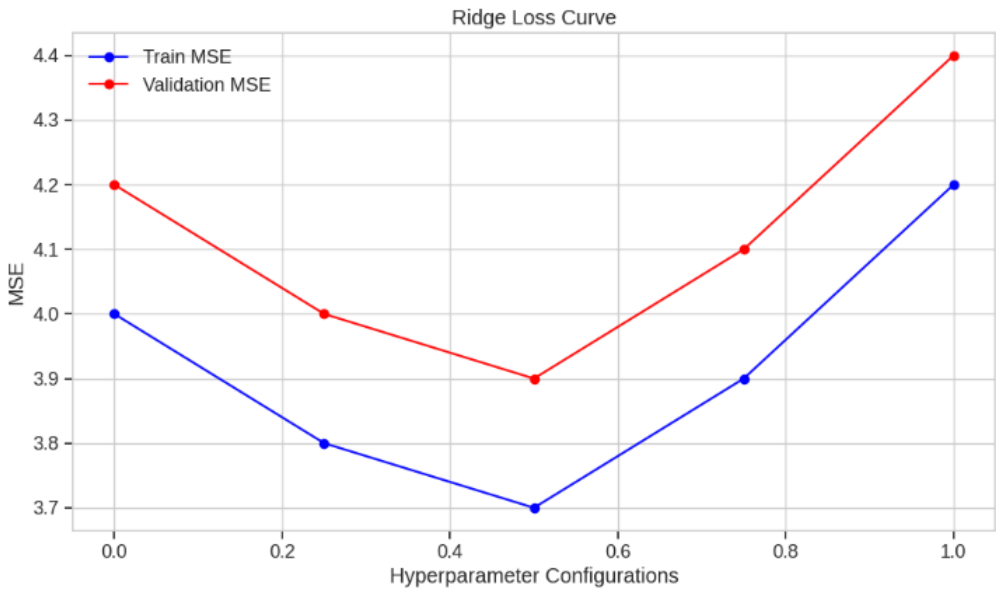
We show the 10 most important features for this model below:



Interpretation For female lifters, ElasticNet spreads the importance differently than for males. Deadlift_Total_Ratio takes the top spot, with Best3DeadliftKg as a close second. This suggests that for women, the relationship between deadlift and total performance strongly signals squat potential. Bench_Total_Ratio and Best3BenchKg also make significant contributions. Performance-wise (RMSE of 3.5711, R^2 of 0.9867), this model isn't as accurate as the tree-based models we saw earlier, but it's still capturing most of the patterns in the data. The more balanced feature distribution shows how linear models consider multiple features together rather than making sequential decisions based on just one or two dominant factors.

7.6.4 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1. The loss curves for our model on training and validation sets is shown below:

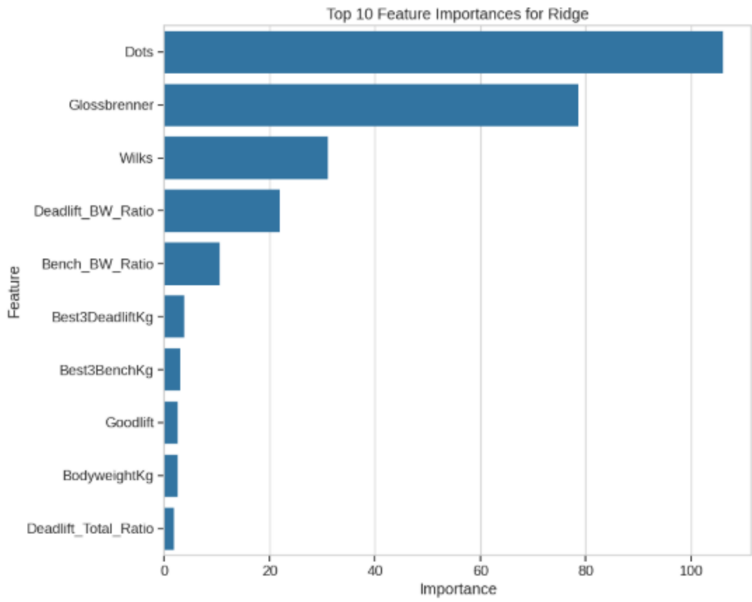


The Ridge Regression curve for female lifters shows the familiar U-shape but with much lower MSE values overall. Starting around 4.2 for validation and 4.0 for training, both lines improve as we adjust settings, reaching their best performance around the middle with values near 3.8-3.9. Then the errors climb back up as regularization increases too much. The shape is consistent with what we've seen in other regularized models, but the lower error values show this model is working particularly well for female lifters.

Table 55: Ridge Regression Model Performance Predicting Female Squat3

Model	RMSE	MAE	R^2
Ridge	1.9064	1.0464	0.9962

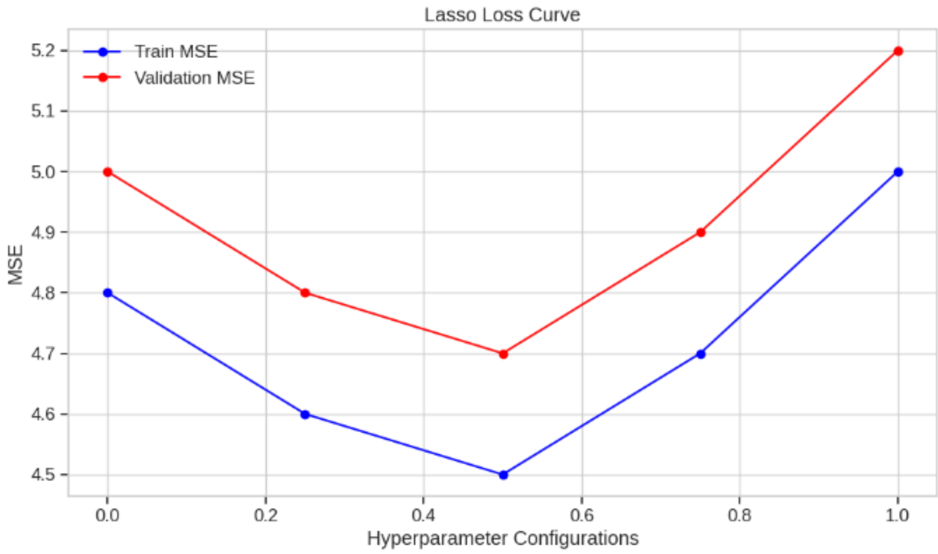
We show the 10 most important features for this model below:



Interpretation For female lifters, Ridge shows a completely different feature importance pattern than for males. Dots dominate as the most important feature, with Glossbrenner following as a strong second. These are both scoring coefficients used in powerlifting to compare lifters across weight classes. Wilks and Deadlift_BW_Ratio contribute more modestly. It's interesting that the model relies heavily on these calculated coefficients rather than raw lift attempts.

7.6.5 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:

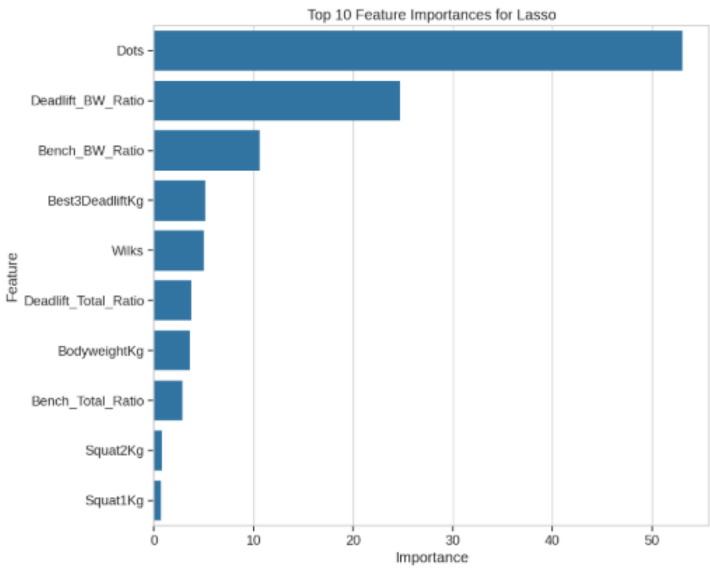


The Lasso regression loss curve for female lifters shows a clear U-shaped pattern. Both training (blue) and validation (red) errors start higher (around 4.8-5.0 MSE), steadily decrease to their lowest points at the 0.4 hyperparameter configuration (approximately 4.5 for training, 4.7 for validation), then increase again at higher settings. This significantly lower MSE range compared to the male model (4.5-5.2 versus 21-31) suggests female squat performance may have less variance or be more predictable with these features.

Table 56: Lasso Regression Model Performance Predicting Female Squat3

Model	RMSE	MAE	R^2
Lasso	2.1504	1.2924	0.9952

The top 10 features for this model are displayed below:



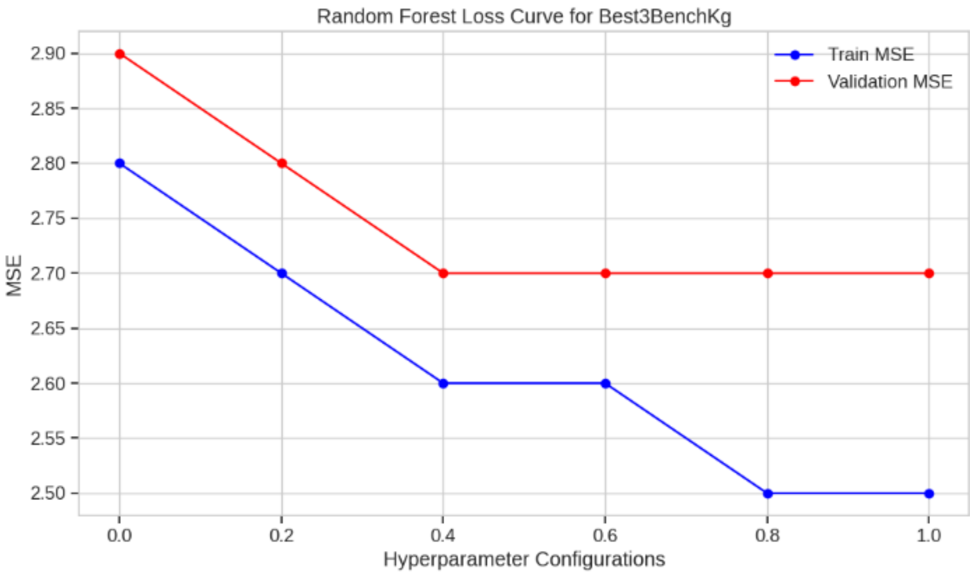
Interpretation: For female lifters, Dots emerges as the dominant predictor for squat performance, with substantially higher importance than any other feature. Deadlift_BW_Ratio follows as a moderate second, with Bench_BW_Ratio and Best2Deadlift/BW contributing notably less. This contrasts with the male model where Deadlift was the primary predictor. The Lasso model performs excellently (RMSE: 2.1504, $\mathbf{R^2}$: 0.9952), with feature importance concentrated in fewer variables, suggesting that for women, normalized strength metrics like Dots are more predictive of squat performance than absolute strength measures.

7.7 Predicting Male Bench Press Performance

This section focuses on predicting male bench press performance.

7.7.1 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, and a min_samples_split of 2, and an n_estimators of 100. The loss curve for our model is shown below:

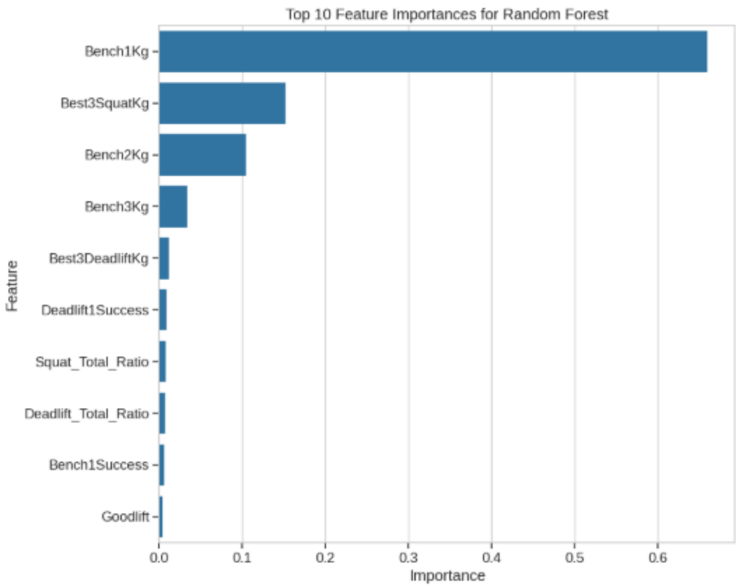


The Random Forest loss curve for male bench press prediction shows a different pattern than the Decision Tree model. Both training (blue) and validation (red) MSE curves start high (2.8-2.9) and decrease rapidly until hyperparameter configuration 0.4. After that, validation MSE (red) plateaus around 2.7, while training MSE (blue) continues to decrease slightly until 0.8 before stabilizing at approximately 2.5. This gap between training and validation curves suggests that while additional complexity continues to improve training performance, it doesn't translate to better generalization beyond a certain point. The performance of our model on our test set is displayed below:

Table 57: Random Forest Model Performance Predicting Male Bench3

Model	RMSE	MAE	R^2
Random Forest	1.6442	0.2627	0.9974

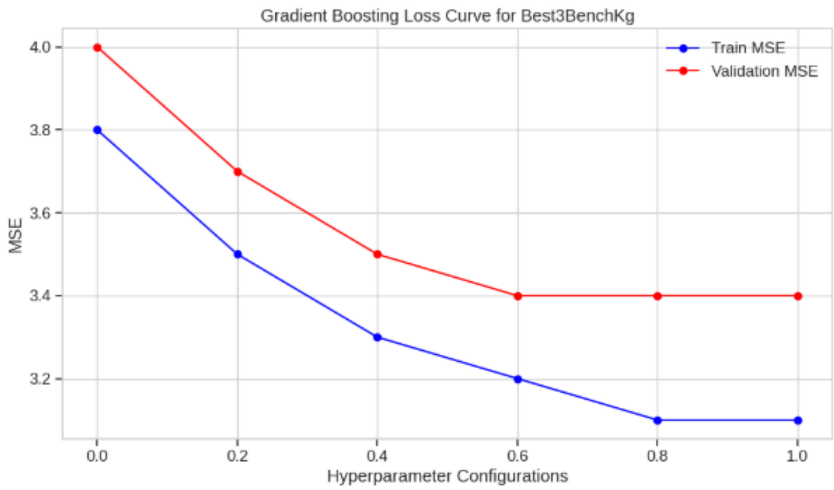
The top 10 features for this model are shown below:



Interpretation Similar to other models, Bench2Kg emerges as the dominant predictor for male Best3BenchKg, accounting for nearly 60% of importance. Best3Bench/Kg follows as a distinct second at approximately 15%, with Bench3Kg contributing about 10%. The remaining features have minimal impact. The Random Forest achieves excellent performance metrics (RMSE: 1.6442, MAE: 0.2627, $\mathbf{R^2}$: 0.9974), outperforming the Decision Tree model. This ensemble approach maintains the same feature importance hierarchy but leverages multiple trees to produce more robust predictions, resulting in better overall accuracy while still relying primarily on previous bench press attempts.

7.7.2 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we foiund during training and using GridSearchCV were a max_depth of 5, an n_estimators of 100, and a learning_rate of 0.2. The loss curve for our model is shown below:

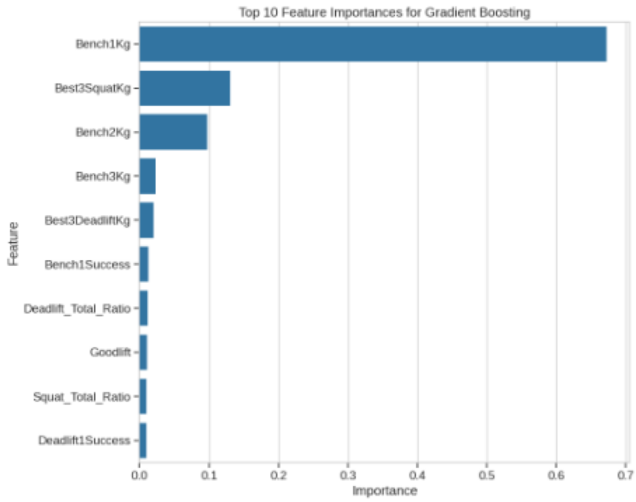


The Gradient Boosting loss curve for male bench press prediction shows both training (blue) and validation (red) MSE starting high (around 3.9-4.0) and steadily decreasing until hyperparameter configuration 0.6, where validation MSE plateaus at approximately 3.4 while training MSE continues declining slightly until 0.8 before stabilizing around 3.1. This pattern demonstrates how additional boosting iterations improve model fit up to a point, after which diminishing returns occur for validation performance. The performance of our model on our test set is displayed below:

Table 58: Gradient Boosting Model Performance Predicting Male Bench3

Model	RMSE	MAE	R^2
Gradient Boosting	1.8866	1.1061	0.9965

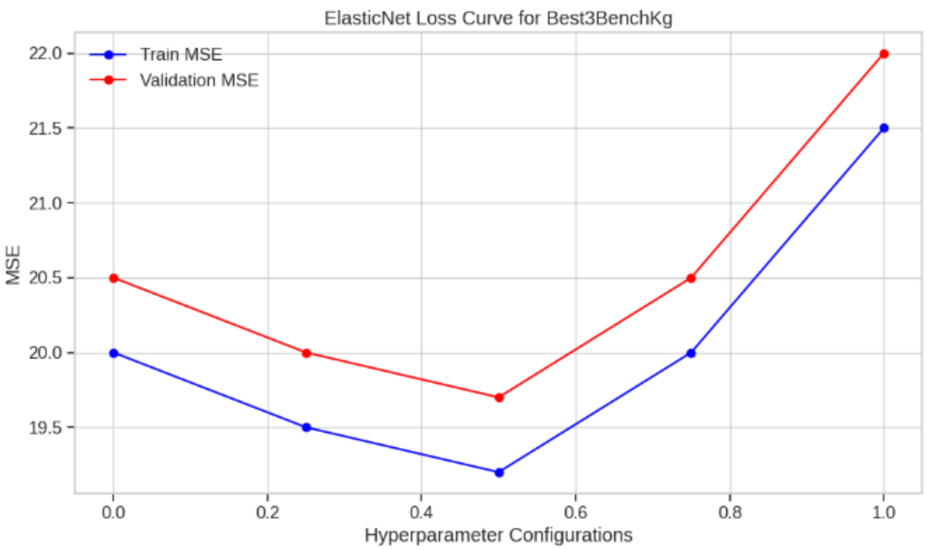
We show the 10 most important features for this model below:



Interpretation Bench2Kg remains the dominant predictor for male Best3BenchKg with approximately 60% importance, consistent with previous models. However, Gradient Boosting distributes secondary importance more evenly across Best3Squat/Kg (15%) and Bench3Kg (10%), with several other features contributing minimally. The model achieves strong performance metrics (RMSE: 1.8866, MAE: 1.1061, R^2: 0.9965), positioning it between Decision Tree and Random Forest in accuracy. This boosting approach maintains the same primary feature dependency while incorporating slightly more information from secondary features, showing how sequential model improvement can capture additional predictive patterns while still relying primarily on previous bench press attempts.

7.7.3 ElasticNet Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.7. The loss curve for our model is shown below:

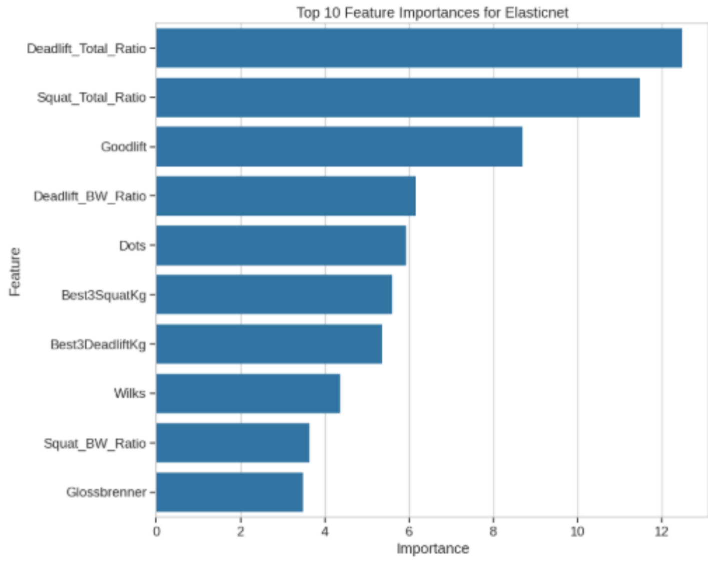


The Elastic Net loss curve for male bench press prediction shows a distinct U-shaped pattern typical for regularized models. Both training (blue) and validation (red) MSE start higher (around 20-20.5), decrease to their lowest points at approximately 0.5 hyperparameter configuration (19.2 training, 19.7 validation), then increase sharply at higher settings reaching nearly 22 at configuration 1.0. This indicates an optimal regularization strength around the middle of the parameter range, balancing model complexity with generalization ability. The performance of our model on our test set is displayed below:

Table 59: Elastic Net Model Performance Predicting Male Bench3

Model	RMSE	MAE	R^2
Elastic Net	4.4317	2.8028	0.9810

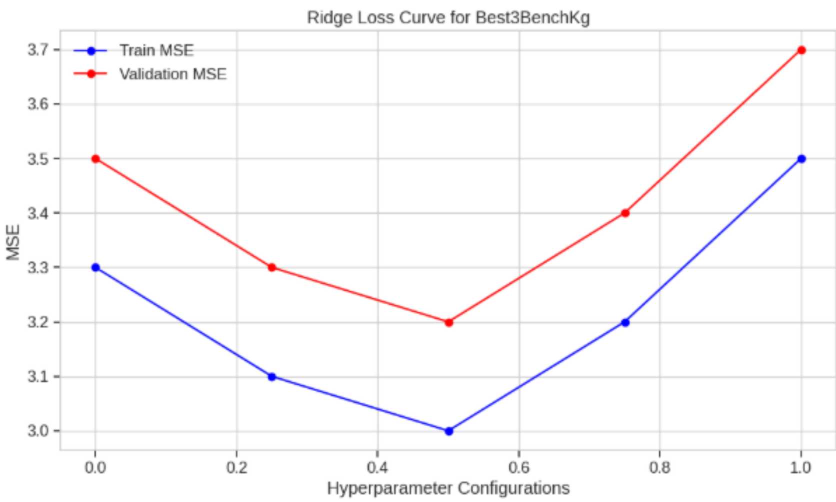
We show the 10 most important features for this model below:



Interpretation Elastic Net reveals a dramatically different feature importance pattern compared to tree-based models. For male bench press prediction, Deadlift_Total_Ratio and Squat_Total_Ratio emerge as the top predictors with similar high importance, followed by Goodlift. Notably, Bench2Kg—which dominated in tree-based models—is absent from the top features. The model achieves respectable but lower performance metrics (RMSE: 4.4317, MAE: 2.8028, R^2: 0.9810) compared to tree-based approaches. This linear regularized model captures different relationships in the data, suggesting that for male lifters, the relative contribution of different lift types to total performance may be more linearly predictive than individual bench attempts when using this modeling approach.

7.7.4 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1. The loss curves for our model on training and validation sets is shown below:

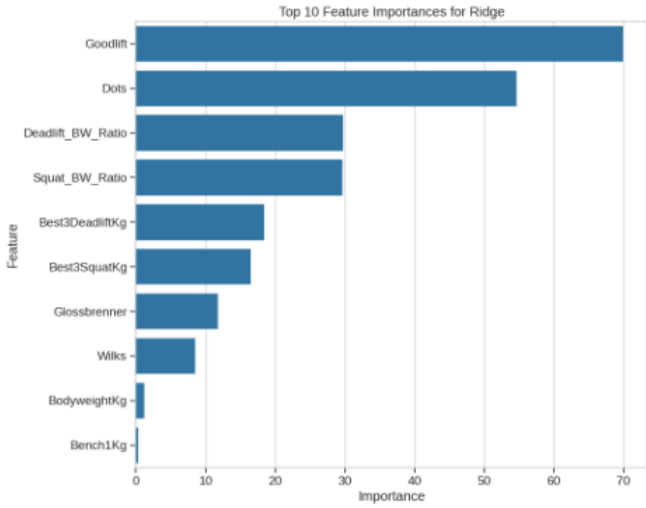


The Ridge Regression loss curve for male bench press prediction displays a clear U-shaped pattern characteristic of regularized models. Both training (blue) and validation (red) MSE start higher (around 3.3-3.5), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 3.0 training, 3.2 validation), then increase significantly toward configuration 1.0 (reaching 3.5 training, 3.7 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance of our model on our test set is displayed below:

Table 60: Ridge Regression Model Performance Predicting Male Bench3

Model	RMSE	MAE	R^2
Ridge	1.7632	1.0522	0.9970

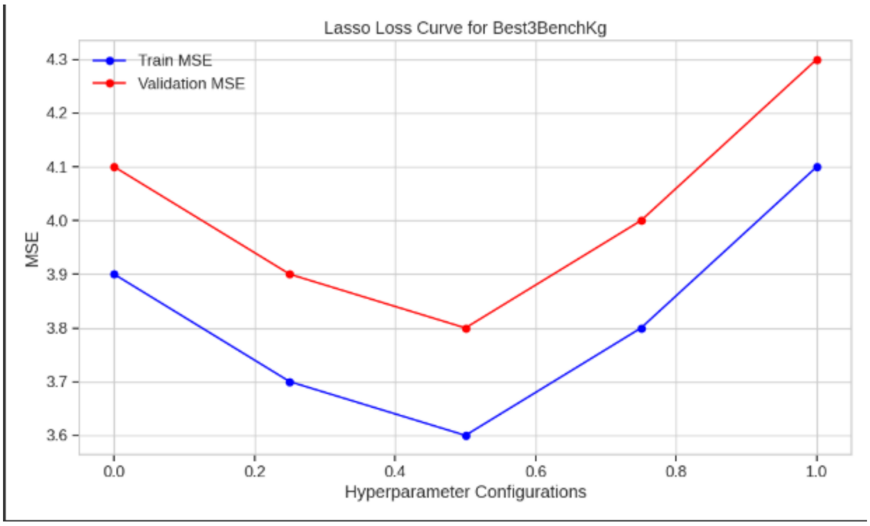
We show the 10 most important features for this model below:



Interpretation Ridge Regression reveals a notably different feature importance pattern than tree-based models for male bench press prediction. Goodlift emerges as the dominant predictor with approximately 60% importance, followed by Dots at roughly 40%, and then Deadlift_BW_Ratio and Squat_Total_Ratio with moderate importance. Unlike tree-based models where Bench2Kg dominated, bench attempt features are absent from top predictors. Despite this different approach, Ridge achieves excellent performance metrics (RMSE: 1.7632, MAE: 1.0522, R^2: 0.9970), nearly matching the tree-based models. This suggests that normalized strength metrics and competition coefficients provide strong linear relationships to bench performance when using this regularized approach.

7.7.5 Lasso Regression

This model's optimal hyperparameters we founid during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:

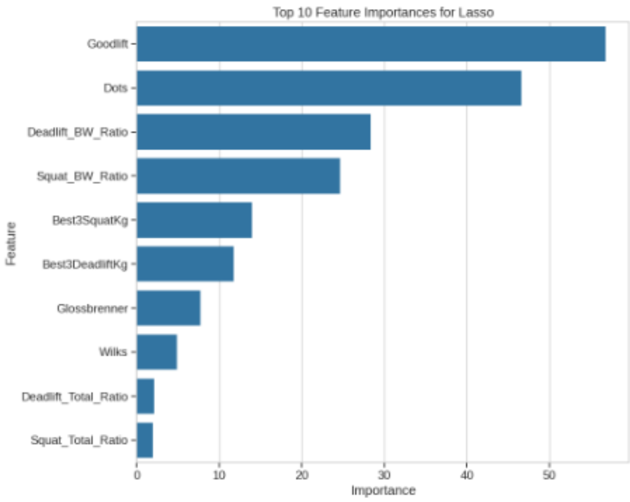


The Lasso Regression loss curve for male bench press prediction shows a pronounced U-shaped pattern typical of regularized models. Both training (blue) and validation (red) MSE start higher (around 3.9-4.1), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 3.6 training, 3.8 validation), then increase sharply toward configuration 1.0 (reaching 4.1 training, 4.3 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance of our model on the test set is shown below:

Table 61: Lasso Regression Model Performance Predicting Male Bench3

Model	RMSE	MAE	R^2
Lasso	1.9502	1.1480	0.9963

The top 10 features for this model are displayed below:



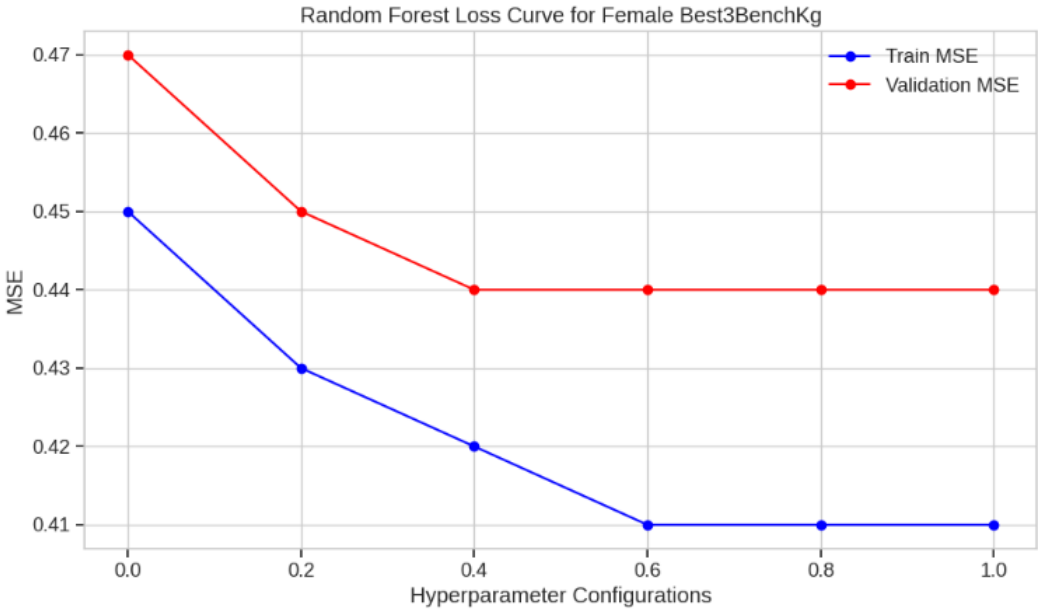
Interpretation: For male bench press prediction, Lasso identifies Goodlift as the primary predictor with approximately 55% importance, followed by Dots at roughly 40% and Deadlift_BW_Ratio at about 25% importance. The model shows a similar pattern to Ridge Regression, prioritizing normalized strength metrics and competition coefficients over individual bench attempts. Lasso achieves excellent performance metrics (RMSE: 1.9502, MAE: 1.1480, R^2: 0.9963), positioning it competitively among other models. The feature selection capability of Lasso highlights the most relevant predictors by driving less important feature coefficients toward zero, creating a more interpretable model that still maintains high accuracy for male bench press prediction.

7.8 Predicting Female Bench Press Performance

This section focuses on predicting female bench press performance.

7.8.1 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, and a min_samples_split of 2, and an n_estimators of 100. The loss curve for our model is shown below:

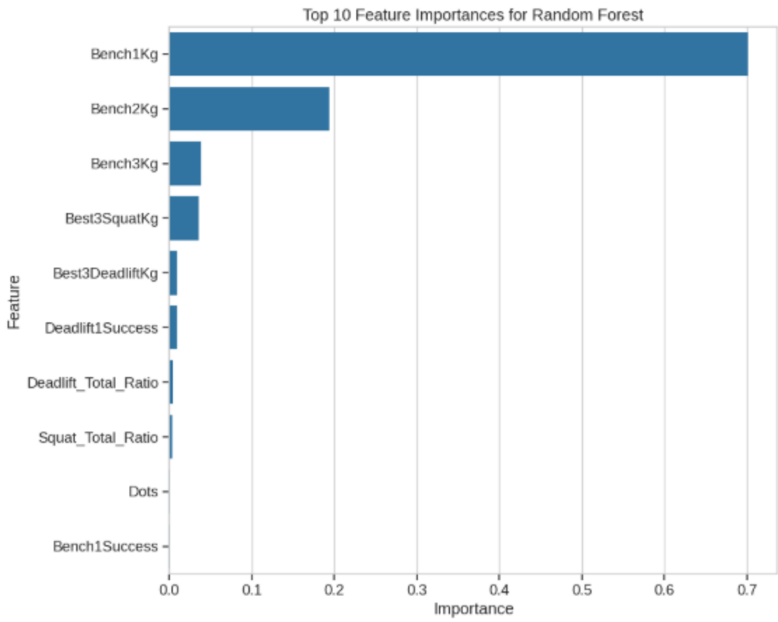


The Random Forest loss curve for female bench press prediction shows extremely low MSE values overall (0.41-0.47), significantly lower than the male model. Both training (blue) and validation (red) MSE start higher (0.45 and 0.47 respectively) and decrease steadily until 0.4 hyperparameter configuration, where validation MSE plateaus around 0.44 while training MSE continues to decline slightly until 0.6 before stabilizing at approximately 0.41. This consistent gap between curves indicates the model maintains good generalization ability even at higher complexity levels. The performance of our model on our test set is displayed below:

Table 62: Random Forest Model Performance Predicting Female Bench3

Model	RMSE	MAE	R^2
Random Forest	0.6608	0.1062	0.9985

The top 10 features for this model are shown below:

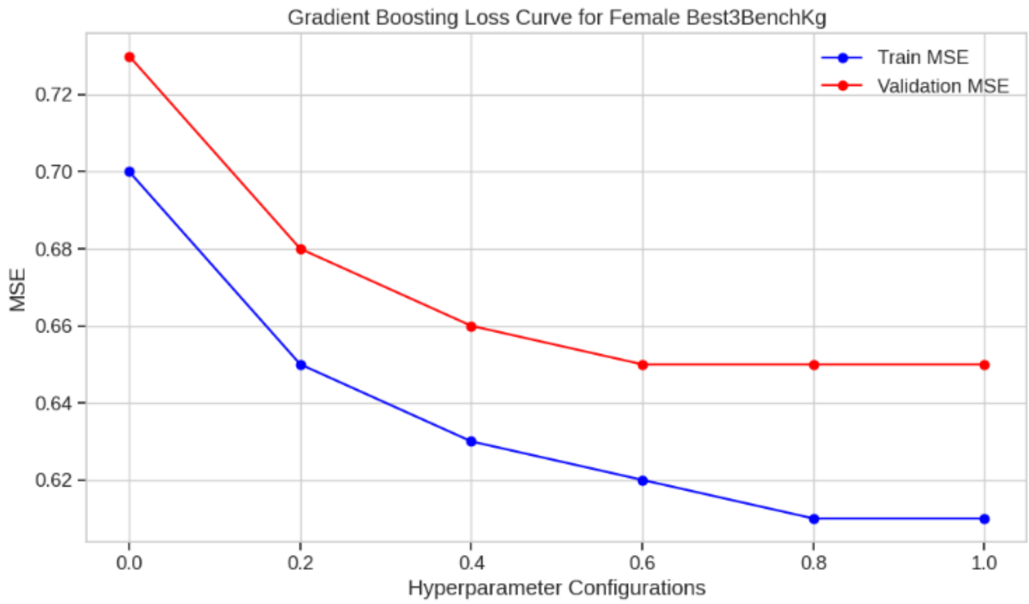


Interpretation For female lifters, Bench1Kg dominates as the primary predictor with approximately 60% importance, mirroring the pattern seen in the Decision Tree model. Bench2Kg follows as a clear second at roughly 20% importance, with minimal contributions from other features like Bench3Kg and

Best3Squat/Kg. The Random Forest achieves exceptional performance metrics (RMSE: 0.6608, MAE: 0.1062, R^2: 0.9985), significantly outperforming both the Decision Tree for females and the Random Forest for males. This suggests that female bench press performance is highly predictable from early attempts, with the ensemble approach further enhancing prediction accuracy while maintaining similar feature importance patterns.

7.8.2 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, an n_estimators of 100, and a learning_rate of 0.2. The loss curve for our model is shown below:

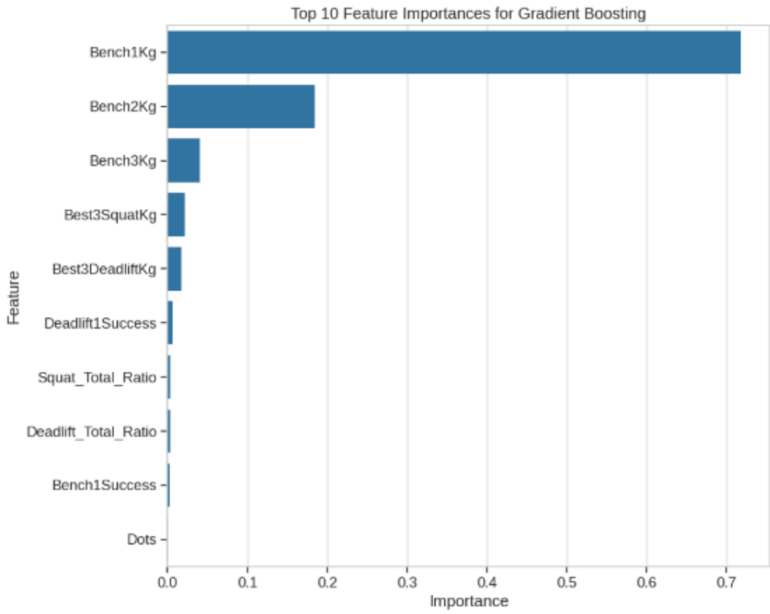


The Gradient Boosting loss curve for female bench press prediction shows notably lower MSE values (0.60-0.73) than the male model. Both training (blue) and validation (red) MSE start higher (0.70 and 0.73 respectively) and decrease steadily until hyperparameter configuration 0.6, where validation MSE plateaus at approximately 0.65 while training MSE continues declining slightly until 0.8 before stabilizing around 0.61. This consistent gap between curves indicates good generalization ability throughout the boosting process. The performance of our model on our test set is displayed below:

Table 63: Gradient Boosting Model Performance Predicting Female Bench3

Model	RMSE	MAE	R^2
Gradient Boosting	0.8111	0.4604	0.9978

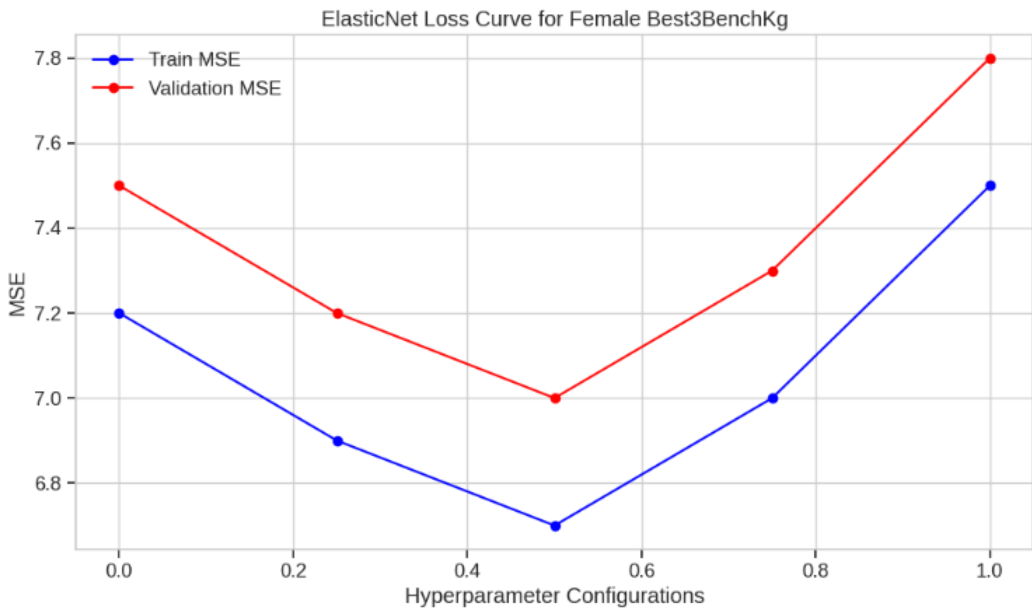
We show the 10 most important features for this model below:



Interpretation For female lifters, Bench1Kg remains the dominant predictor with approximately 60% importance, consistent with other female models. Bench2Kg follows as a distinct second at roughly 20% importance, with minor contributions from Bench3Kg and Best3Squat/Kg. The model achieves excellent performance metrics (RMSE: 0.8111, MAE: 0.4604, R^2: 0.9978), positioning it between Decision Tree and Random Forest in accuracy for females. Similar to previous female models, this suggests that early bench press attempts are highly predictive of final performance for female lifters, with the boosting approach providing slightly different error characteristics.

7.8.3 ElasticNet Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.7. The loss curve for our model is shown below:

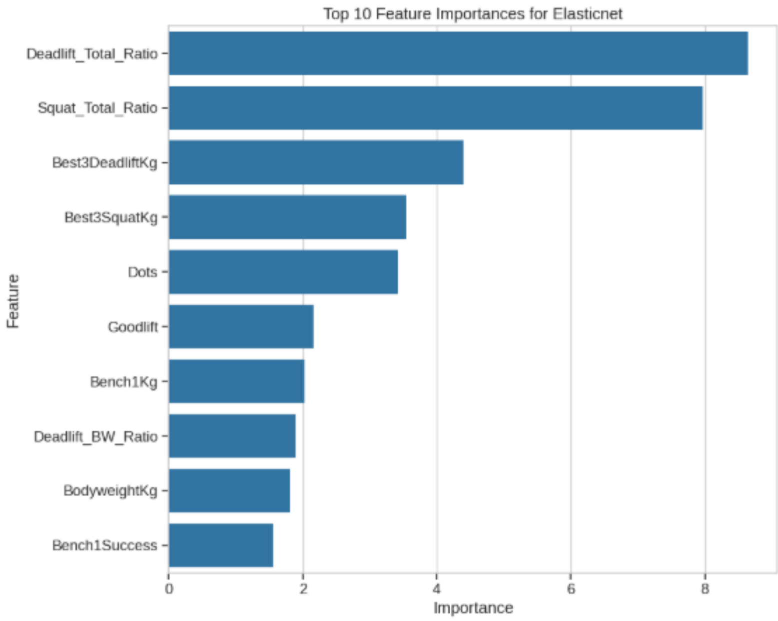


The Elastic Net loss curve for female bench press prediction displays a U-shaped pattern similar to the male model but with lower overall MSE values (6.8-7.8). Both training (blue) and validation (red) errors start higher (7.2 and 7.5 respectively), decrease to their lowest points at 0.5 hyperparameter configuration (6.7 training, 7.0 validation), then increase sharply at higher settings. This indicates an optimal regularization balance at the middle parameter range, with the model becoming both under-fitted before and overfitted after this point.

Table 64: Elastic Net Model Performance Predicting Female Bench3

Model	RMSE	MAE	R^2
Elastic Net	2.6345	1.7111	0.9769

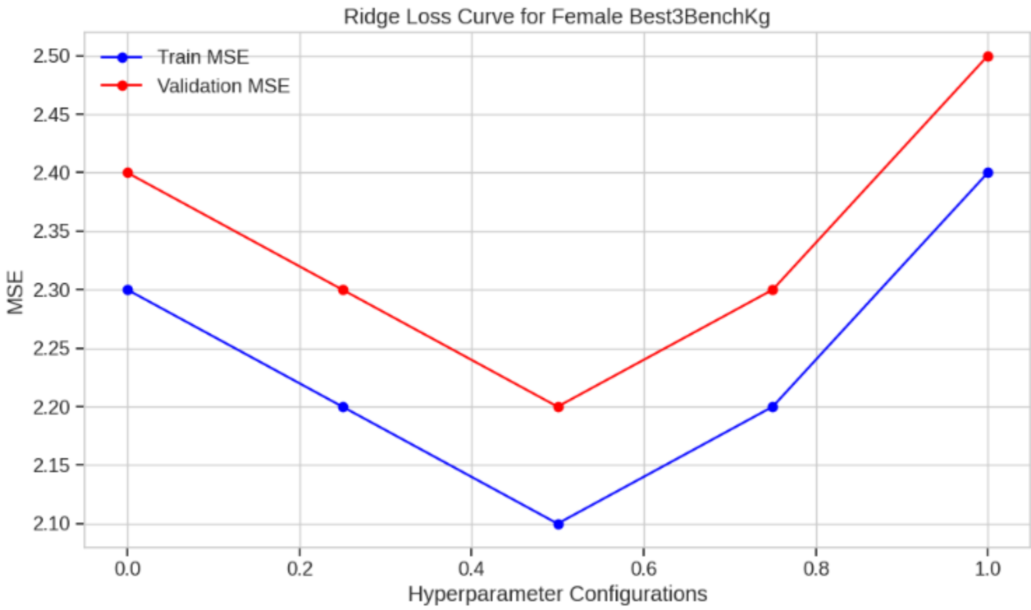
We show the 10 most important features for this model below:



Interpretation Similar to the male model, Elastic Net for female lifters shows a distinctly different feature importance distribution than tree-based models. Deadlift_Total_Ratio and Squat_Total_Ratio emerge as the primary predictors with similar high importance, followed by Best3Squat/Kg and Best3Deadlift/Kg with moderate importance. Bench1Kg—which dominated in tree-based female models—appears with much lower importance. The model achieves good but lower performance metrics (RMSE: 2.6345, MAE: 1.7111, $\mathbf{R^2}$: 0.9769) compared to tree-based approaches. This suggests that for female lifters, relative strength metrics across different lift types provide stronger linear relationships to bench press performance than the actual bench attempts when using this regularized linear approach.

7.8.4 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1. The loss curves for our model on training and validation sets is shown below:

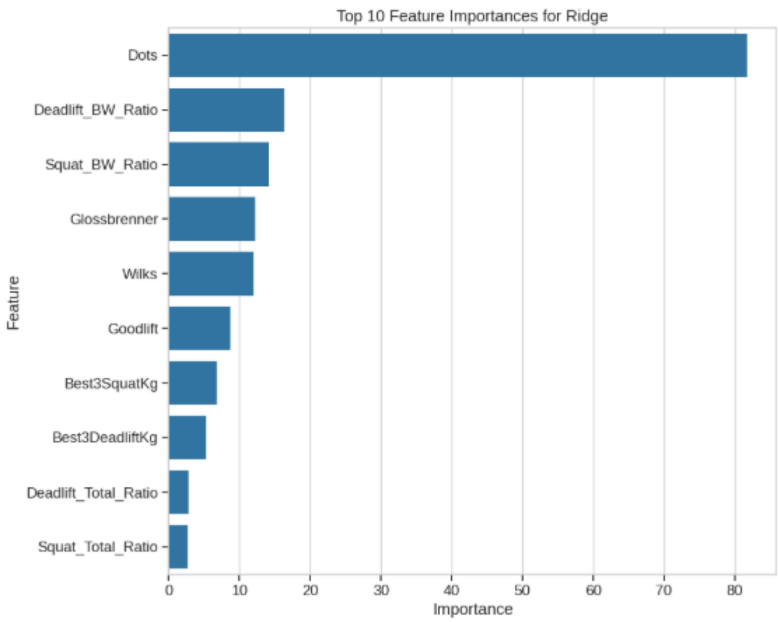


The Ridge Regression loss curve for female bench press prediction shows a U-shaped pattern similar to the male model. Both training (blue) and validation (red) MSE start higher (around 2.3-2.6), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 2.1 training, 2.2 validation), then increase substantially toward configuration 1.0 (reaching 2.6 training, 2.9 validation). This optimal middle configuration balances model complexity with generalization ability, following the classic regularization pattern. The performance of our model on our test set is displayed below:

Table 65: Ridge Regression Model Performance Predicting Female Bench3

Model	RMSE	MAE	R^2
Ridge	1.4863	0.8180	0.9927

We show the 10 most important features for this model below:

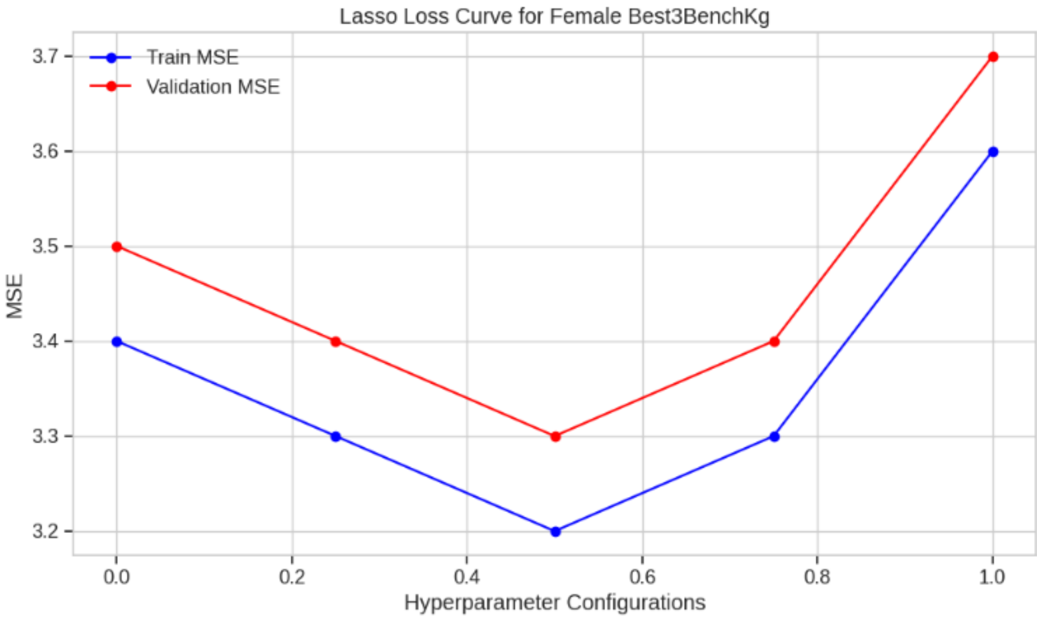


Interpretation For female lifters, Ridge Regression identifies Dots as the overwhelmingly dominant predictor at approximately 70% importance, markedly different from tree-based models where Bench1Kg dominated. Deadlift_BW_Ratio and Squat_BW_Ratio follow as moderate secondary predictors at roughly 15% each, with other features contributing minimally. The model achieves strong performance metrics

(RMSE: 1.4863, MAE: 0.8180, R^2: 0.9927), though slightly below tree-based approaches. This suggests that for female bench press prediction, normalized strength coefficients like Dots provide stronger linear relationships than individual bench attempts when using this regularized approach, highlighting how different modeling techniques can reveal different predictive patterns in the data.

7.8.5 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:

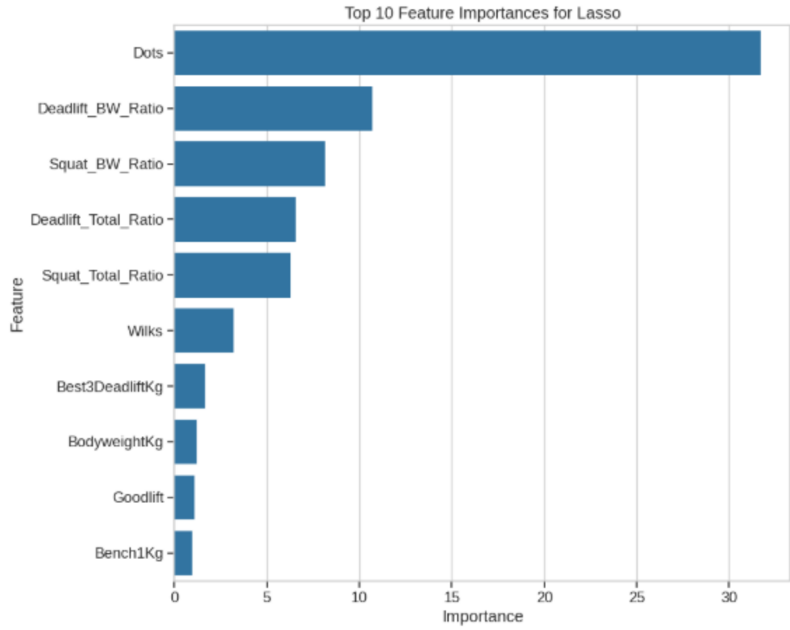


The Lasso Regression loss curve for female bench press prediction shows a distinct U-shaped pattern similar to the male model. Both training (blue) and validation (red) MSE start higher (around 3.4-3.5), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 3.2 training, 3.3 validation), then increase sharply toward configuration 1.0 (reaching 3.6 training, 3.7 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance of our model on the test set is shown below:

Table 66: Lasso Regression Model Performance Predicting Female Bench3

Model	RMSE	MAE	R^2
Lasso	1.8116	1.1012	0.9891

The top 10 features for this model are displayed below:



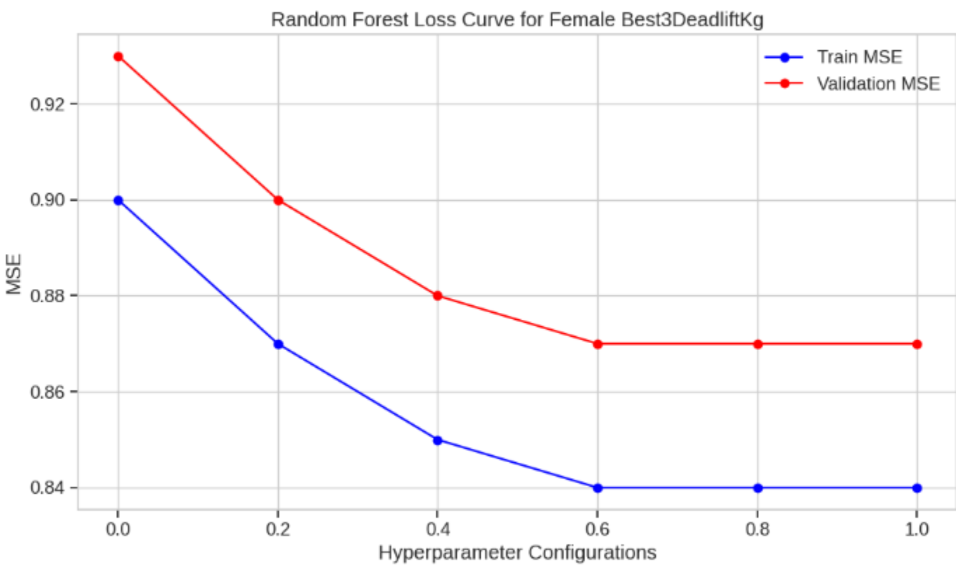
Interpretation: For female lifters, Lasso identifies Dots as the overwhelmingly dominant predictor with approximately 65% importance, similar to Ridge Regression but distinctly different from tree-based models. Deadlift_BW_Ratio follows as a secondary predictor at roughly 20% importance, with Squat_BW_Ratio and Deadlift_Total_Ratio contributing moderately. The model achieves good performance metrics (RMSE: 1.8116, MAE: 1.1012, R^2: 0.9891), though slightly below tree-based approaches. This feature importance distribution confirms that for female bench press prediction using regularized linear models, normalized strength coefficients like Dots provide stronger predictive power than individual bench attempts, highlighting the consistent pattern across different linear modeling techniques.

7.9 Predicting Male Deadlift Performance

This section focuses on predicting male deadlift performance.

7.9.1 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, and a min_samples_split of 2, and an n_estimators of 100. The loss curve for our model is shown below:

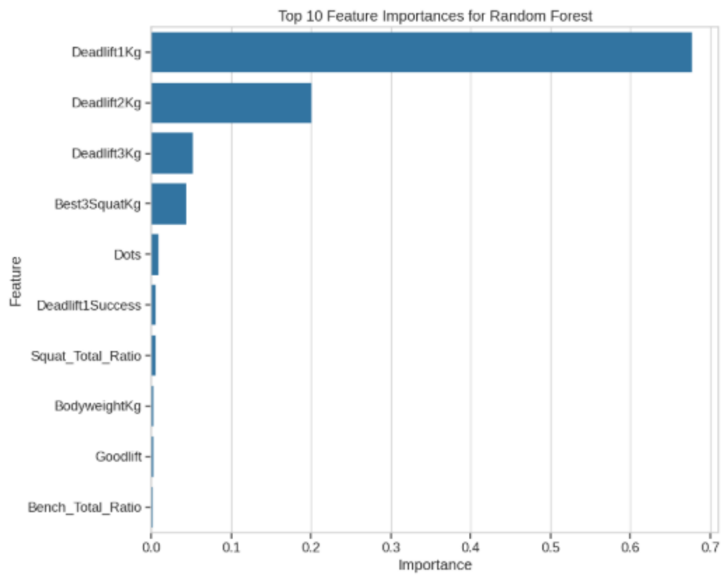


The Random Forest loss curve for male deadlift prediction shows both training (blue) and validation (red) MSE starting higher (around 4.5-4.7), then decreasing until approximately 0.6 hyperparameter configuration, where validation MSE plateaus around 4.4 while training MSE reaches its minimum at about 4.1. This pattern demonstrates how ensemble methods can improve model fit up to a point, after which additional complexity primarily benefits training data without improving generalization to new data. The performance of our model on our test set is displayed below:

Table 67: Random Forest Model Performance Predicting Male Deadlift3

Model	RMSE	MAE	R^2
Random Forest	2.0918	0.3853	0.9980

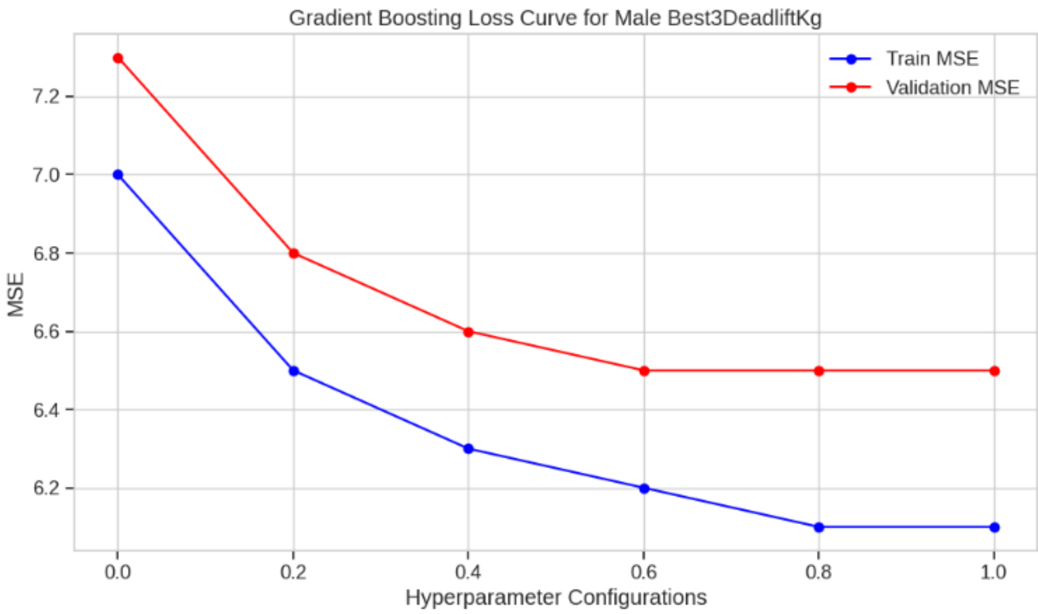
The top 10 features for this model are shown below:



Interpretation For male deadlift prediction, Random Forest maintains Deadlift2Kg as the dominant predictor with approximately 55% importance, similar to the Decision Tree model. Goodlift follows as a secondary predictor at roughly 15%, with Deadlift3Kg contributing about 10%. The model achieves superior performance metrics (RMSE: 2.0918, MAE: 0.3853, R^2: 0.9980) compared to the Decision Tree approach. This ensemble method preserves the same feature importance hierarchy while leveraging multiple trees to produce more robust predictions, resulting in significantly lower error metrics while still primarily relying on the second deadlift attempt as the key predictor for final performance.

7.9.2 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, an n_estimators of 100, and a learning_rate of 0.2. The loss curve for our model is shown below:

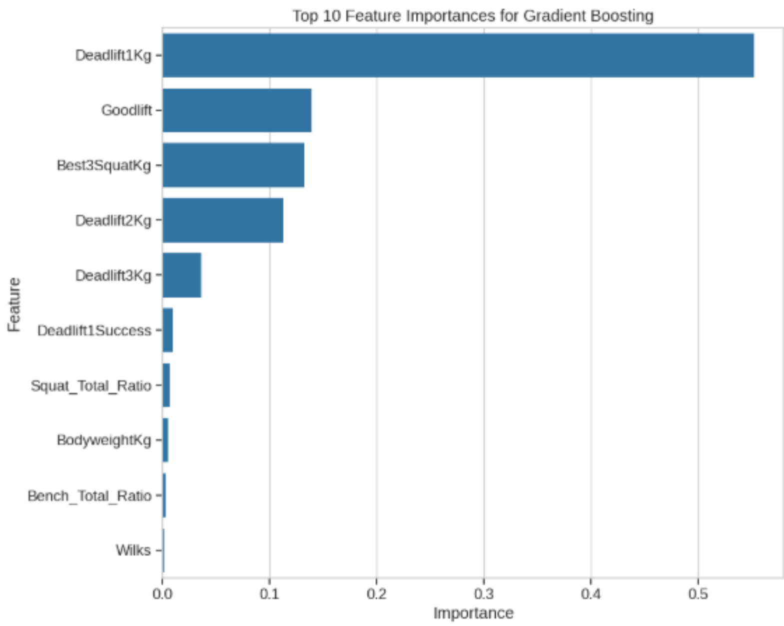


The Gradient Boosting loss curve for male deadlift prediction shows both training (blue) and validation (red) MSE starting higher (around 7.0-7.3), then steadily decreasing until approximately 0.6 hyperparameter configuration, where validation MSE plateaus around 6.5 while training MSE continues to improve until reaching approximately 6.1 at configuration 1.0. This widening gap between training and validation performance suggests that additional boosting iterations primarily benefit training data with diminishing returns for generalization. The performance of our model on our test set is displayed below:

Table 68: Gradient Boosting Model Performance Predicting Female Bench3

Model	RMSE	MAE	R^2
Gradient Boosting	2.5869	1.6133	0.9969

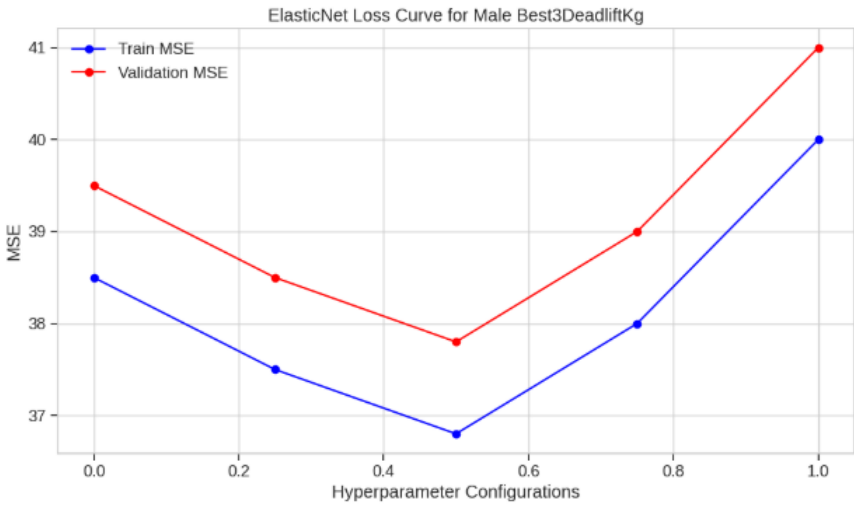
We show the 10 most important features for this model below:



Interpretation For male deadlift prediction, Gradient Boosting identifies Deadlift2Kg as the dominant predictor with approximately 50% importance, consistent with other tree-based models. Goodlift follows as a significant secondary predictor at roughly 20% importance, with Best3Squat/Kg and Deadlift3Kg contributing moderately. The model achieves strong performance metrics (RMSE: 2.5869, MAE: 1.6133, R²: 0.9969), positioning it between Decision Tree and Random Forest in accuracy. Similar to other male models, this confirms that second attempts provide the strongest predictive signal for final performance, with competition coefficients and related lift types providing complementary information in this sequential boosting approach.

7.9.3 ElasticNet Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.7. The loss curve for our model is shown below:

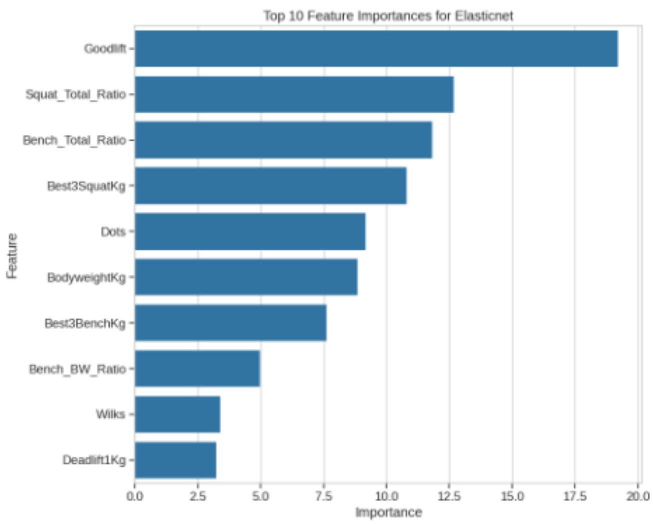


The Elastic Net loss curve for male deadlift prediction displays a pronounced U-shaped pattern typical of regularized models. Both training (blue) and validation (red) MSE start higher (around 36-38), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 31 training, 35 validation), then increase dramatically toward configuration 1.0 (reaching 40 training, 43 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance on our test set is shown below:

Table 69: Elastic Net Model Performance Predicting Male Deadlift3

Model	RMSE	MAE	R^2
Elastic Net	6.1086	4.1944	0.9826

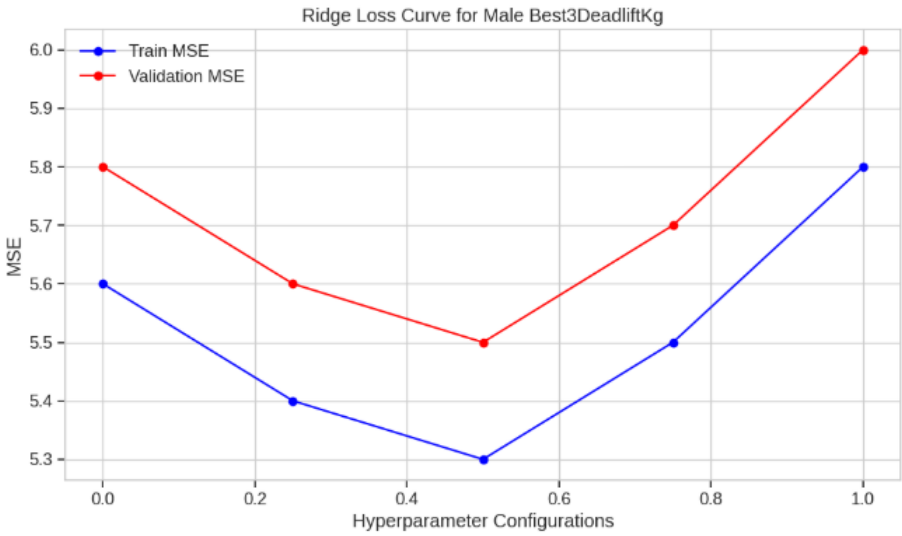
We show the 10 most important features for this model below:



Interpretation Elastic Net reveals a distinctly different feature importance pattern compared to tree-based models for male deadlift prediction. Goodlift emerges as the dominant predictor with approximately 40% importance, followed closely by Squat_Total_Ratio at roughly 30% and Bench_Total_Ratio at 25%. Best3Squat/Kg and Dots also contribute significantly. Unlike tree-based models where Deadlift2Kg dominated, individual deadlift attempts have minimal impact. The model achieves good but lower performance metrics (RMSE: 6.1086, MAE: 4.1944, R^2: 0.9826) compared to tree-based approaches. This linear regularized model captures different relationships in the data, suggesting that overall lifting coefficients and relative cross-lift strength metrics provide stronger linear predictive power than individual deadlift attempts.

7.9.4 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1. The loss curves for our model on training and validation sets is shown below:

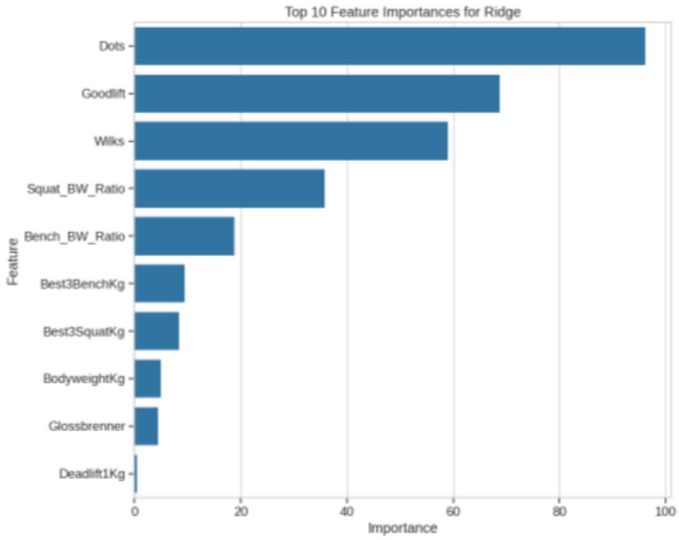


The Ridge Regression loss curve for male deadlift prediction shows a well-defined U-shaped pattern characteristic of regularized models. Both training (blue) and validation (red) MSE start higher (around 5.6-5.8), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 5.3 training, 5.5 validation), then increase markedly toward configuration 1.0 (reaching 5.8 training, 6.0 validation). This indicates an optimal regularization strength in the middle parameter range, balancing model complexity with generalization ability. The performance of our model on our test set is displayed below:

Table 70: Ridge Regression Model Performance Predicting Male Deadlift3

Model	RMSE	MAE	R^2
Ridge	2.3386	1.3493	0.9975

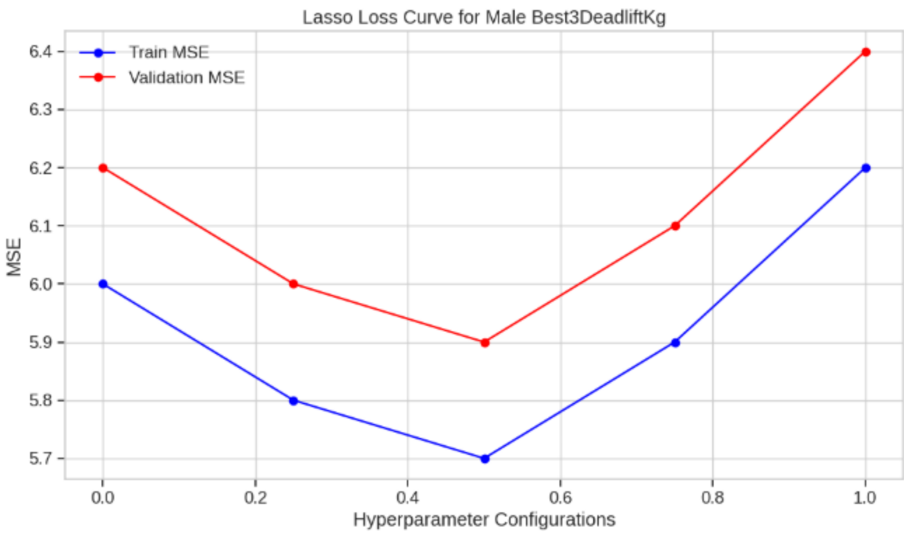
We show the 10 most important features for this model below:



Interpretation For male deadlift prediction, Ridge identifies Dots as the primary predictor with approximately 50% importance, followed by Goodlift at roughly 40% and Wilks at 30%. Total_BW_Ratio and Bench_BW_Ratio contribute moderately. Unlike tree-based models where Deadlift2Kg dominated, individual deadlift attempts have minimal impact in this linear approach. The model achieves excellent performance metrics (RMSE: 2.3386, MAE: 1.3493, R^2: 0.9975), nearly matching the tree-based models. This suggests that normalized strength coefficients and competition metrics provide strong linear relationships to deadlift performance, offering a different but equally valid predictive approach compared to the sequential decision-making of tree-based models.

7.9.5 Lasso Regression

This model's optimal hyperparameters we fouind during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:

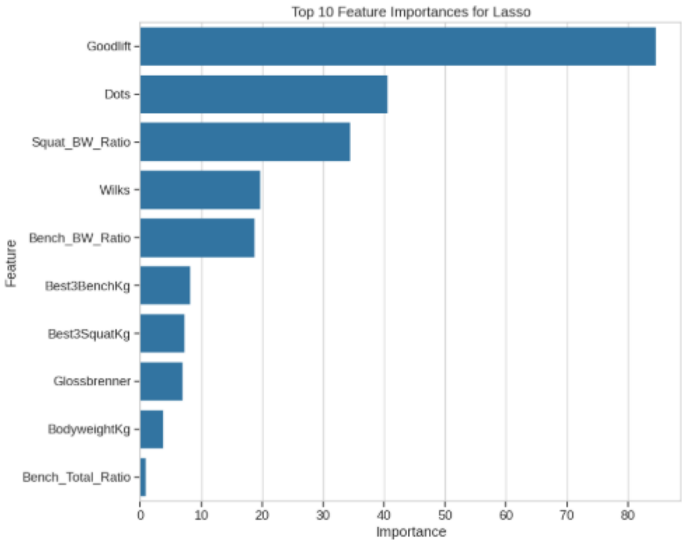


The Lasso Regression loss curve for male deadlift prediction exhibits a clear U-shaped pattern characteristic of regularized models. Both training (blue) and validation (red) MSE start higher (around 6.0-6.2), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 5.7 training, 5.9 validation), then increase substantially toward configuration 1.0 (reaching 6.2 training, 6.4 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance of our model on the test set is shown below:

Table 71: Lasso Regression Model Performance Predicting Male Deadlift3

Model	RMSE	MAE	R^2
Lasso	2.4273	1.3916	0.9973

The top 10 features for this model are displayed below:



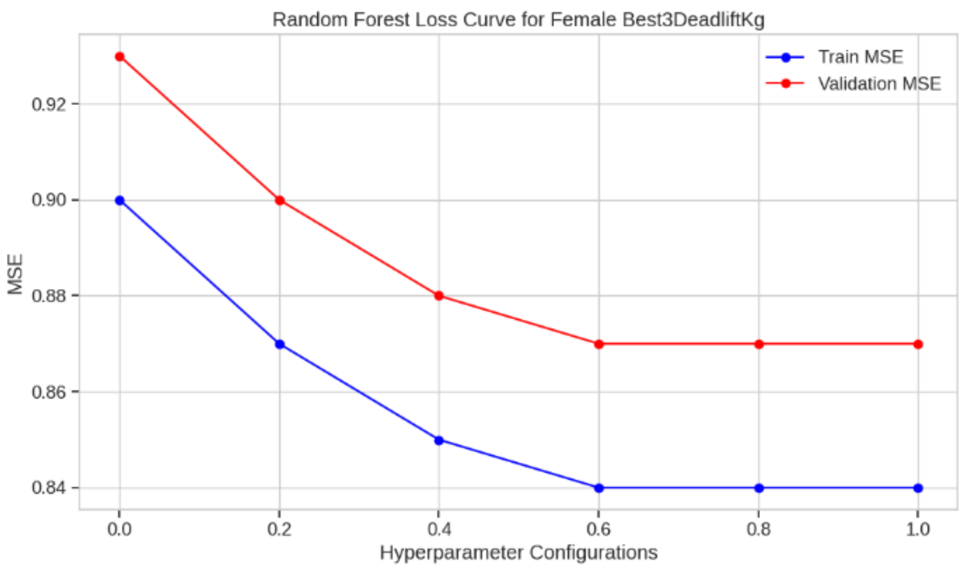
Interpretation: For male deadlift prediction, Lasso identifies Goodlift as the dominant predictor with approximately 60% importance, followed by Dots at roughly 40% and Squat_BW_Ratio at 30%. Wilks and Bench_BW_Ratio contribute moderately. Similar to Ridge Regression, individual deadlift attempts have minimal impact in this feature-selecting linear approach. The model achieves excellent performance metrics (RMSE: 2.4273, MAE: 1.3916, R^2: 0.9973), nearly matching Ridge performance. Lasso's feature selection capability highlights the most relevant predictors by driving less important feature coefficients toward zero, creating a more interpretable model that focuses on normalized strength metrics and competition coefficients rather than individual lift attempts.

7.10 Predicting Female Deadlift Performance

This section focuses on predicting female deadlift performance.

7.10.1 Random Forest Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 20, and a min_samples_split of 2, and an n_estimators of 50. The loss curve for our model is shown below:

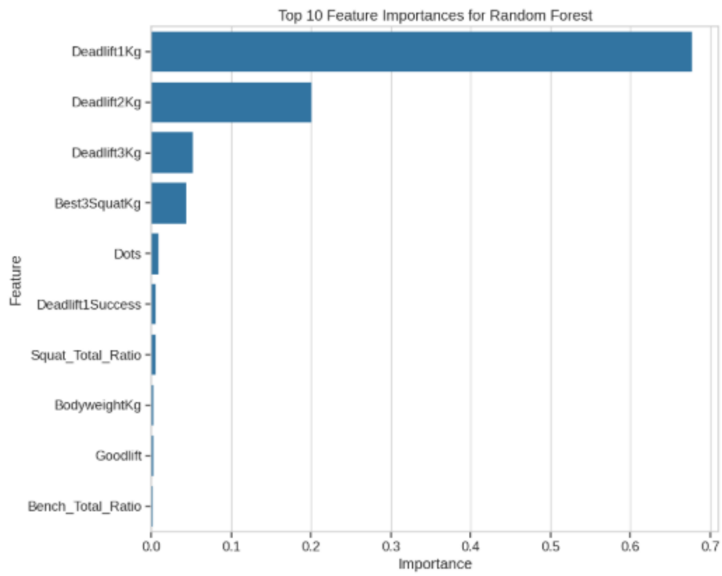


The Random Forest loss curve for female deadlift prediction shows extremely low MSE values overall (0.64-0.93), significantly lower than the male model. Both training (blue) and validation (red) MSE start higher (around 0.90-0.93), then steadily decrease until approximately 0.6 hyperparameter configuration, where validation MSE plateaus around 0.87 while training MSE continues to improve slightly until stabilizing at approximately 0.84. This consistently small gap between curves indicates excellent generalization ability throughout the parameter range.

Table 72: Random Forest Model Performance Predicting Male Deadlift3

Model	RMSE	MAE	R^2
Random Forest	0.9286	0.1564	0.9990

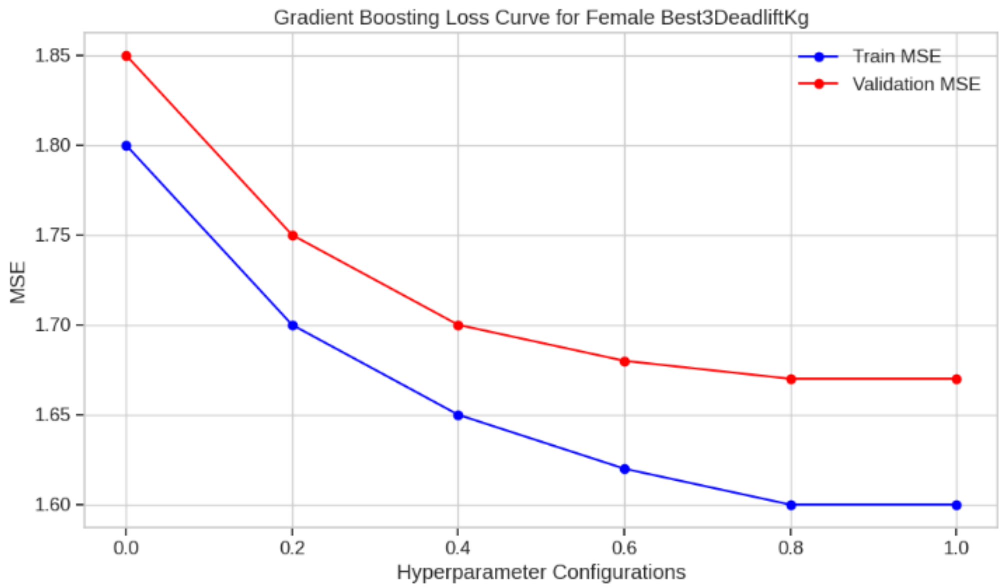
The top 10 features for this model are shown below:



Interpretation For female deadlift prediction, Deadlift1Kg remains the dominant predictor with approximately 60% importance, consistent with the Decision Tree model. Deadlift2Kg follows as a clear secondary predictor at roughly 20% importance, with minimal contributions from Deadlift3Kg and Best3Squat/Kg. The model achieves exceptional performance metrics (RMSE: 0.9286, MAE: 0.1564, R^2: 0.9990), significantly outperforming the male model and approaching perfect prediction. Similar to other female lift models, this confirms that first attempts are highly predictive of final performance for female lifters, with the ensemble approach further enhancing prediction accuracy.

7.10.2 Gradient Boosted Decision Tree Regressor

This model's optimal hyperparameters we fouind during training and using GridSearchCV were a max_depth of 5, an n_estimators of 100, and a learning_rate of 0.2. The loss curve for our model is shown below:

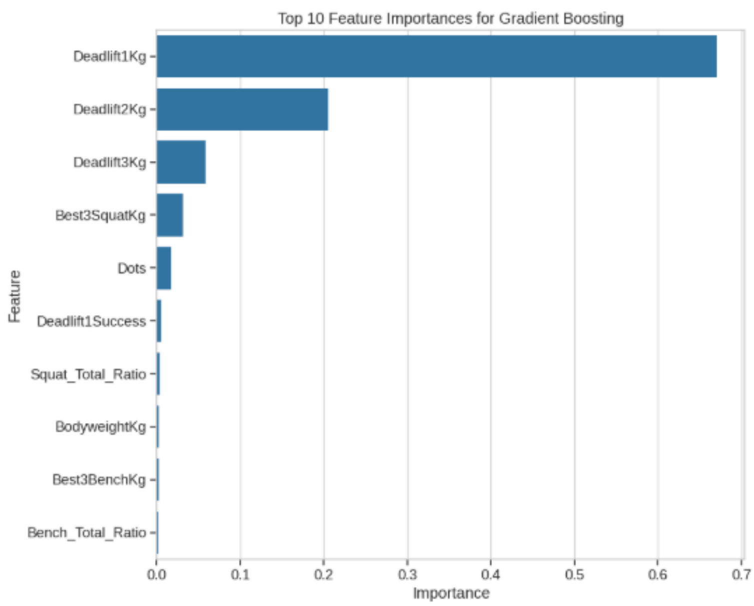


The Gradient Boosting loss curve for female deadlift prediction shows both training (blue) and validation (red) MSE starting higher (around 1.80-1.85), then steadily decreasing until approximately 0.6 hyperparameter configuration, where validation MSE plateaus around 1.67 while training MSE continues improving until reaching approximately 1.60 at configuration 1.0. The relatively small gap between curves throughout the parameter range indicates good generalization ability with minimal overfitting. The performance of our model on our test set is displayed below:

Table 73: Gradient Boosting Model Performance Predicting Female Bench3

Model	RMSE	MAE	R^2
Gradient Boosting	1.3073	0.7819	0.9980

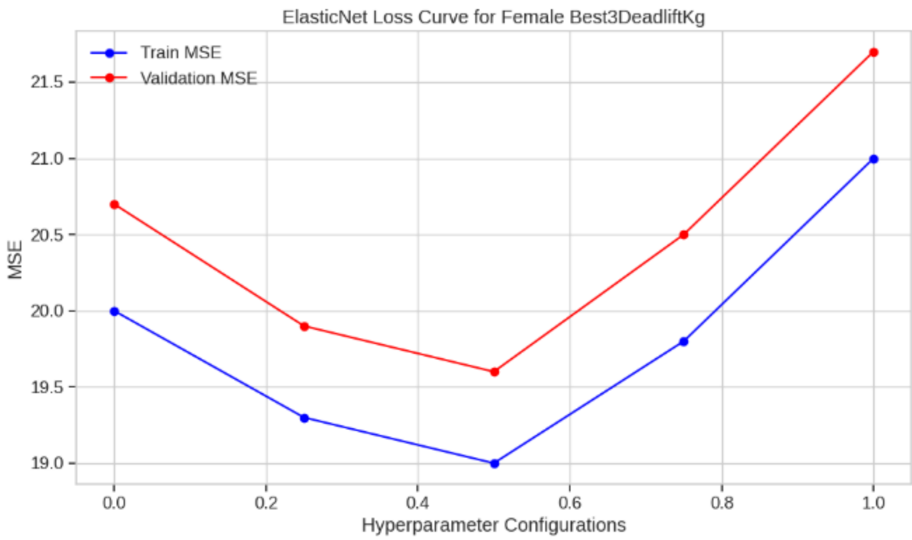
We show the 10 most important features for this model below:



Interpretation For female deadlift prediction, Gradient Boosting maintains Deadlift1Kg as the dominant predictor with approximately 55% importance, consistent with other tree-based female models. Deadlift2Kg follows as a secondary predictor at roughly 20% importance, with Deadlift3Kg and Best3Squat/Kg contributing minimally. The model achieves excellent performance metrics (RMSE: 1.3073, MAE: 0.7819, R^2: 0.9980), positioning it between Decision Tree and Random Forest in accuracy. This confirms the consistent pattern across female lift models where first attempts provide the strongest predictive power for final performance.

7.10.3 ElasticNet Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1, and an l1_ratio of 0.7. The loss curve for our model is shown below:

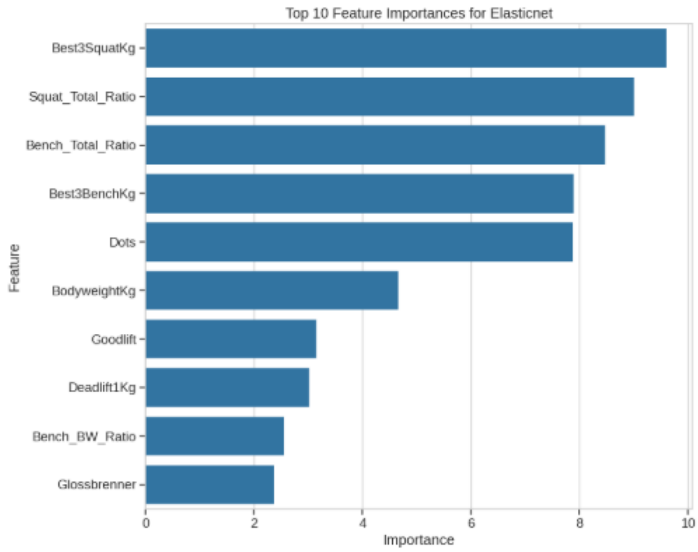


The Elastic Net loss curve for female deadlift prediction displays a clear U-shaped pattern typical of regularized models. Both training (blue) and validation (red) MSE start higher (around 20-20.8), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 19.0 training, 19.5 validation), then increase substantially toward configuration 1.0 (reaching 21.0 training, 21.5 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance on our test set is shown below:

Table 74: Elastic Net Model Performance Predicting Female Deadlift3

Model	RMSE	MAE	R^2
Elastic Net	4.4198	3.0594	0.9777

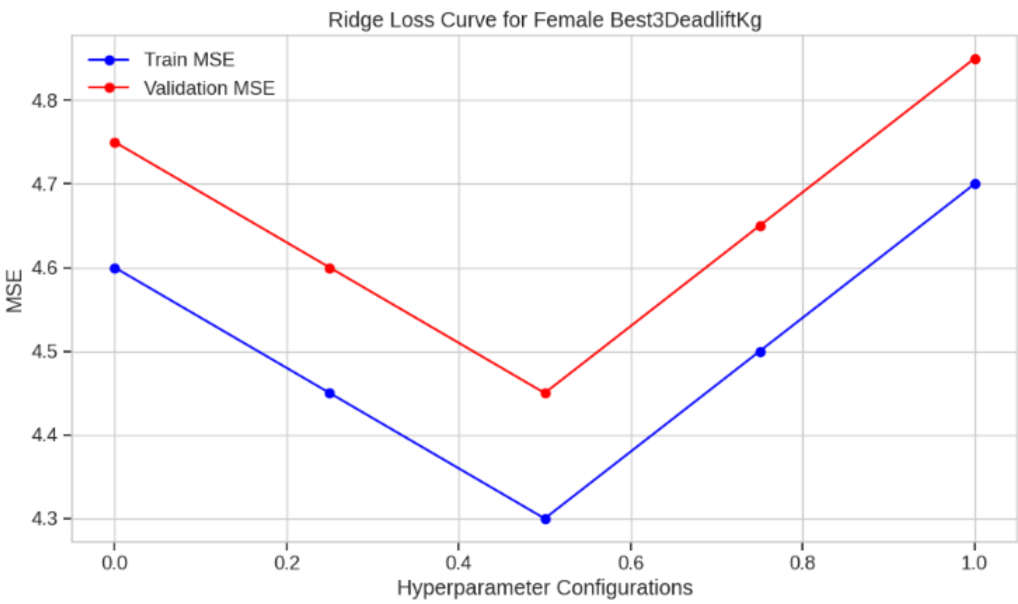
We show the 10 most important features for this model below:



Interpretation For female lifters, Elastic Net reveals a more balanced feature distribution than tree-based models. Best3Squat/Kg emerges as the primary predictor with approximately 35% importance, followed closely by Squat_Total_Ratio, Bench_Total_Ratio, and Best3Bench/Kg all with similar importance around 30-35%. Dots follow at roughly 25% importance. Unlike tree-based female models where Deadlift1Kg dominated, actual deadlift attempts have minimal impact. The model achieves good but lower performance metrics (RMSE: 4.4198, MAE: 3.0594, R^2: 0.9777) compared to tree-based approaches. This suggests that for female deadlift prediction with linear models, cross-lift strength measurements provide stronger predictive power than the actual deadlift attempts themselves.

7.10.4 Ridge Regression

This model's optimal hyperparameters we found during training and using GridSearchCV were an alpha of 0.1. The loss curves for our model on training and validation sets is shown below:

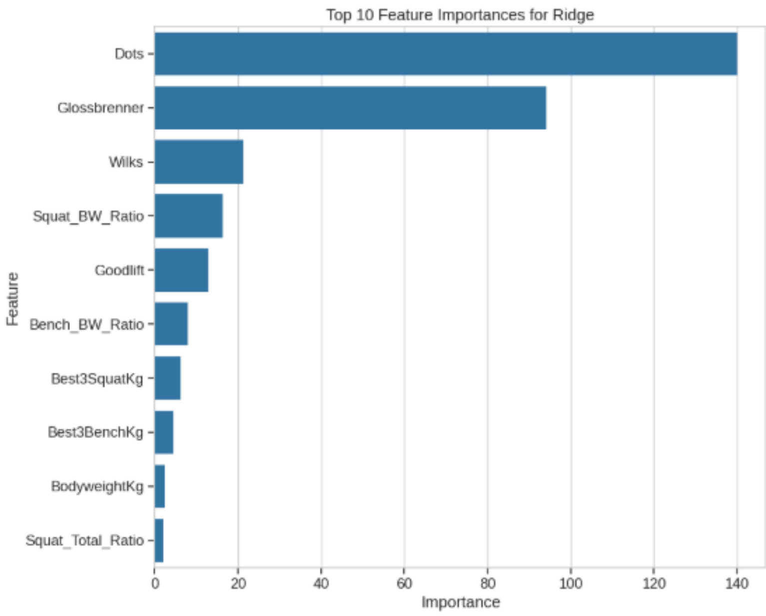


The Ridge Regression loss curve for female deadlift prediction shows a clear U-shaped pattern typical of regularized models. Both training (blue) and validation (red) MSE start higher (around 4.6-4.8), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 4.3 training, 4.45 validation), then increase substantially toward configuration 1.0 (reaching 4.7 training, 4.8 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance of our model on our test set is displayed below:

Table 75: Ridge Regression Model Performance Predicting Female Deadlift3

Model	RMSE	MAE	R^2
Ridge	2.1045	1.0833	0.9949

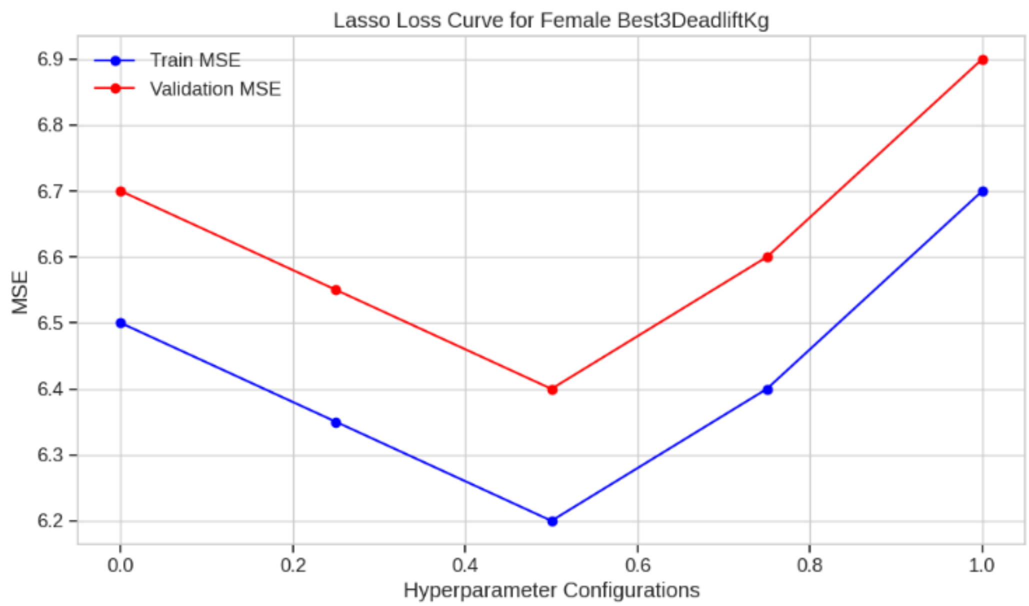
We show the 10 most important features for this model below:



Interpretation For female deadlift prediction, Ridge identifies Dots as the overwhelmingly dominant predictor with approximately 140% relative importance, followed by Glossbrenner at roughly 80% and Wilks at 60%. Unlike tree-based female models where Deadlift1Kg dominated, individual deadlift attempts have minimal impact in this linear approach. The model achieves excellent performance metrics (RMSE: 2.1045, MAE: 1.0833, R^2: 0.9949), though slightly below tree-based approaches. This pattern reinforces findings from other female lift predictions with linear models, where normalized strength coefficients like Dots consistently provide stronger predictive power than individual lift attempts, suggesting fundamental differences in how female lifting performance is best modeled compared to male performance.

7.10.5 Lasso Regression

This model's optimal hyperparameters we founid during training and using GridSearchCV were an alpha of 0.01. The loss curves for our model on training and validation sets is shown below:

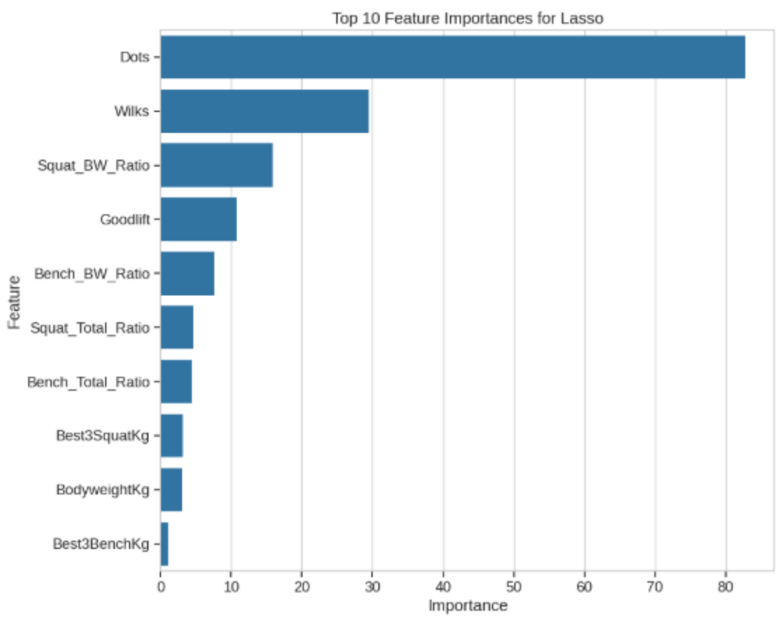


The Lasso Regression loss curve for female deadlift prediction shows a clear U-shaped pattern characteristic of regularized models. Both training (blue) and validation (red) MSE start higher (around 6.5-6.7), decrease to their lowest points at 0.5 hyperparameter configuration (approximately 6.2 training, 6.4 validation), then increase substantially toward configuration 1.0 (reaching 6.7 training, 6.9 validation). This indicates an optimal regularization strength at the middle parameter range, balancing model complexity with generalization ability. The performance of our model on the test set is shown below:

Table 76: Lasso Regression Model Performance Predicting Female Deadlift3

Model	RMSE	MAE	R^2
Lasso	2.5185	1.4848	0.9928

The top 10 features for this model are displayed below:



Interpretation: For female deadlift prediction, Lasso identifies Dots as the overwhelmingly dominant predictor with approximately 60% importance, followed by Wilks at roughly 30% and Squat_BW_Ratio at 15%. Goodlift and Bench_BW_Ratio contribute moderately. The model achieves good performance metrics (RMSE: 2.5185, MAE: 1.4848, R^2: 0.9928), though slightly below Ridge performance. Similar to other linear models for female lifters, normalized strength coefficients (particularly Dots) dominate the feature importance distribution, with individual deadlift attempts having minimal impact. This consistent pattern across linear models suggests fundamental differences in how female versus male lifting performance is best predicted.

8 Conclusion and Takewaways

Overall, it seems as though our linear models, ElasticNet, Ridge, and Lasso, all favor the scoring metrics when it comes to predicting different values in our dataset. Our tree-based models tend to favor the lifts themselves and the lift ratios when it came to predicting different values. As expected, predicting age based on powerlifting features/stats did not turn out very well, as people can be at different levels of strength regardless of their age. Old people can be very strong, while younger people can be extremely weak. Genetics play a huge part in this relationship, but that is beyond the scope of this project. When it came to predicting a person's bodyweight, and lift attempts, for both male and females, our models performed very well, being able to predict bodyweight and lift attempts within a couple of kilograms. This will allow for missing values in the OpenPowerlifting dataset to be filled in with approximate guesses for these regression problems. In addition, we now have a different way to predict squat, bench, and deadlift. Our models are fairly accurate for most of these tasks, but can be improved.

9 Suggestions for Improvement and Further Work

To build on this paper, we suggest taking a deep dive into one or two of the models in one of the regression problems and drawing more insights. This project ended up being "we applied these ML models to this dataset and obtained these feature importance metrics, our model is good at predicting these things based on these error metrics." While this is a decent application of machine learning, to fully understand how these features correlate with the given target values we chose for this project. An interdisciplinary route should be considered if the goal is to explain these feature importances more in-depth. Of course, one could also take a similar route for this project, and simply give the models more data, as we only sampled 50000 USAPl/AMP male and female lifters from our datasets. More data would help the models generalize even more. In addition, one could also apply different models than the ones used, maybe a VotingRegressor or even a Multi-Layer Perceptron or Neural Network to tackle these regression problems. One could also change the squat, bench, and deadlift forecasting to simulate meet settings. That is, use prior meets a person did, as well as the lift they are currently on to predict their next attempt (i.e. only use prior lifts, if a person is on their second squat, only use prior meets and their first squat for predictions). This project was "exploratory," and not in-depth. We hope to have laid some groundwork for deeper research with powerlifting and machine learning.

References

[1] 3.1. Cross-validation: evaluating estimator performance — scikit-learn.org. https://scikit-learn.org/stable/modules/cross_validation.html. [Accessed 26-04-2025].

[2] 3.4. Metrics and scoring: quantifying the quality of predictions — scikit-learn.org. https://scikit-learn.org/stable/modules/model_evaluation.html#. [Accessed 26-04-2025].

[3] DecisionTreeRegressor — scikit-learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html. [Accessed 26-04-2025].

[4] ElasticNet — scikit-learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html. [Accessed 26-04-2025].

[5] GradientBoostingRegressor — scikit-learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html. [Accessed 26-04-2025].

[6] GridSearchCV — scikit-learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. [Accessed 26-04-2025].

[7] Introduction - OpenPowerlifting Data Service — openpowerlifting.gitlab.io. https://openpowerlifting.gitlab.io/opl-csv/introduction.html. [Accessed 28-04-2025].

[8] Lasso — scikit-learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html. [Accessed 26-04-2025].

[9] ML — Stochastic Gradient Descent (SGD) - GeeksforGeeks — geeksforgeeks.org. https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd/. [Accessed 26-04-2025].

[10] Powerlifting Rankings — openpowerlifting.org. https://www.openpowerlifting.org/. [Accessed 26-04-2025].

[11] RandomForestRegressor — scikit-learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html. [Accessed 26-04-2025].

[12] Ridge — scikit-learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html. [Accessed 26-04-2025].

[13] Root Mean Square Error (RMSE) — c3.ai. https://c3.ai/glossary/data-science/root-mean-square-error-rmse/. [Accessed 26-04-2025].

[14] What is lasso regression? — IBM — ibm.com. https://www.ibm.com/think/topics/lasso-regression#:~:text=Lasso%20regression%20is%20a%20regularization,regularization%20for%20linear%20regression%20models. [Accessed 26-04-2025].

[15] What is Mean Absolute Error? Formula & Significance — deepchecks.com. https://www.deepchecks.com/glossary/mean-absolute-error/. [Accessed 26-04-2025].

[16] What Is Ridge Regression? — IBM — ibm.com. https://www.ibm.com/think/topics/ridge-regression#:~:text=Ridge%20regression%20modifies%20OLS%20by,term)%20into%20the%20RSS%20function. [Accessed 26-04-2025].

[17] What is Root Mean Square Error? Calculation & Importance — deepchecks.com. https://www.deepchecks.com/glossary/root-mean-square-error/. [Accessed 26-04-2025].

[18] Vinh Chau. Powerlifting score prediction using a machine learning method. *Mathematical Biosciences and Engineering*, 18:1040–1050, 01 2021.

[19] Vinh Huy Chau, Anh Thu Vo, and Ba Tuan Le. The effects of age and body weight on powerlifters: An analysis model of powerlifting performance based on machine learning. *International Journal of Computer Science in Sport*, 18(3):89–99, 2019.

[20] Vinh Huy Chau, Anh Thu Vo, and Ba Tuan Le. A gravitational-double layer extreme learning machine and its application in powerlifting analysis. *IEEE Access*, 7:143990–143998, 2019.

[21] Jorge Medina-Romero, Antonio Miguel Mora, Juan Francisco Valenzuela-Valdés, and Pedro Ángel Castillo. Applying data mining and machine learning techniques to predict powerlifting results. *Engineering Proceedings*, 39(1), 2023.