

Valparaiso University

ValpoScholar

---

Symposium on Undergraduate Research and  
Creative Expression (SOURCE)

Office of Sponsored and Undergraduate  
Research

---

Spring 5-1-2020

## Optimizing Recovery Conditions in Collegiate Female Soccer Athletes Using Machine Learning

Mark Young  
mark.young@valpo.edu

Follow this and additional works at: <https://scholar.valpo.edu/cus>

---

### Recommended Citation

Young, Mark, "Optimizing Recovery Conditions in Collegiate Female Soccer Athletes Using Machine Learning" (2020). *Symposium on Undergraduate Research and Creative Expression (SOURCE)*. 906. <https://scholar.valpo.edu/cus/906>

This Oral Presentation is brought to you for free and open access by the Office of Sponsored and Undergraduate Research at ValpoScholar. It has been accepted for inclusion in Symposium on Undergraduate Research and Creative Expression (SOURCE) by an authorized administrator of ValpoScholar. For more information, please contact a ValpoScholar staff member at [scholar@valpo.edu](mailto:scholar@valpo.edu).



# Optimizing recovery conditions in female soccer athletes using machine learning

Mark Young



# Background

- VWSO Tracking
- Lack of predictive models
- Lack of external factors
- Importance of proper recovery



## Related Work

- Fitness trackers and improved data collection
- Advanced biometrics
- Heart rate variability
- HRV studies
- Nonlinear models



## Methods: Data

- Data from VWSO (Firstbeat trackers) & NOAA
  - 98 variables
  - 77 not relevant to this research





## Methods: Data

- Data from VWSO (Firstbeat trackers) & NOAA
  - 98 variables
  - 77 not relevant to this research
- Relevant Variables
  - HRV attributes
  - Intensity
  - Miscellaneous

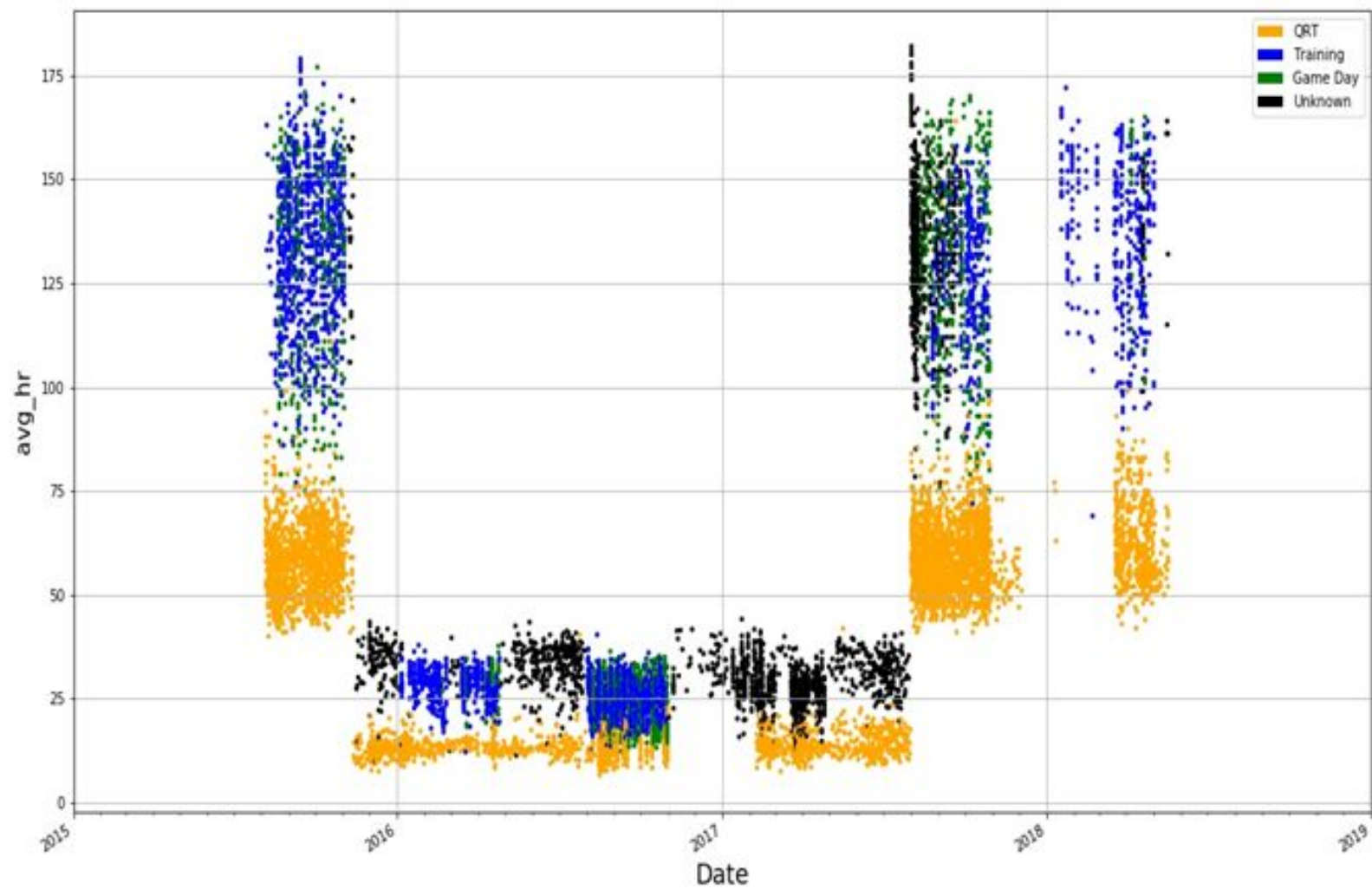




## Methods: Data

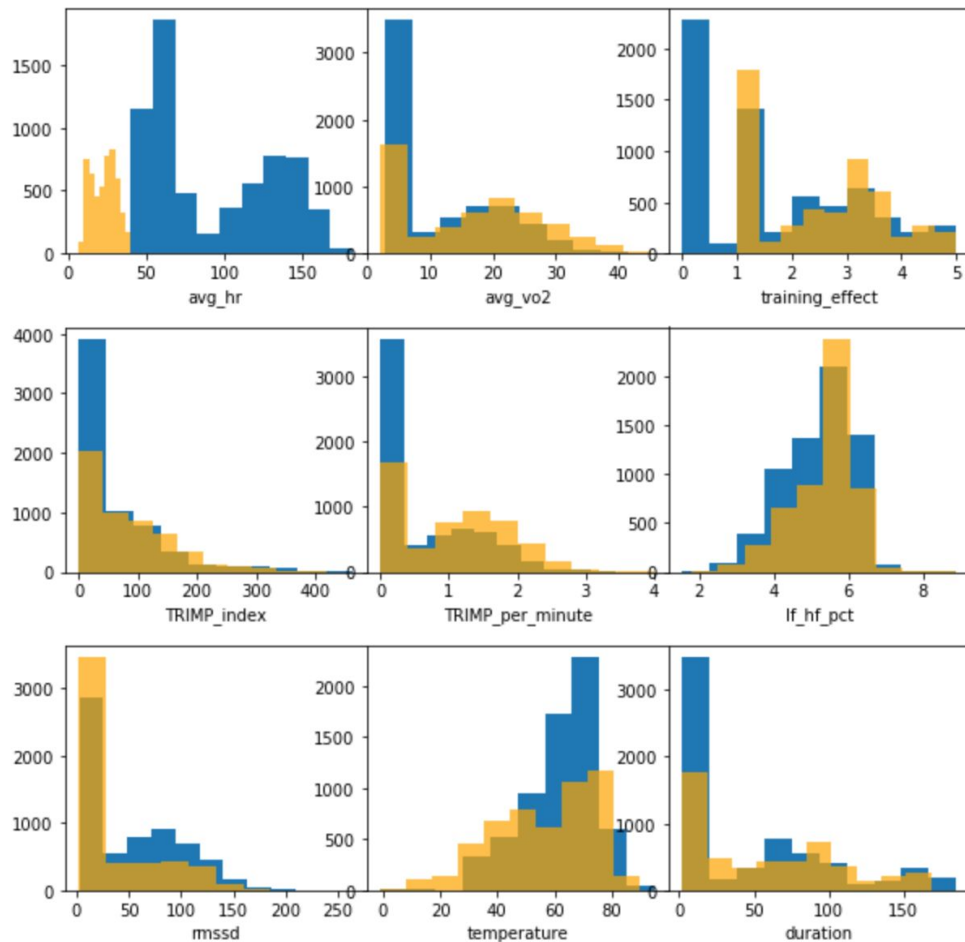
- Data from VWSO (Firstbeat trackers) & NOAA
  - 98 variables
  - 77 not relevant to this research
- Relevant Variables
  - HRV attributes
  - Intensity
  - Miscellaneous
- Data Preprocessing
- Feature Engineering and Transformation







# Distributions of Attributes





## Methods: Data

- Data from VWSO (Firstbeat trackers) & NOAA
  - 98 variables
  - 77 not relevant to this research
- Relevant Variables
  - HRV attributes
  - Intensity
  - Miscellaneous
- Data Preprocessing
- Feature Engineering and Transformation






# Methods: Models

- Team Model and Individual Models
- Linear Mixed Effects Model
- Random Forest Regression



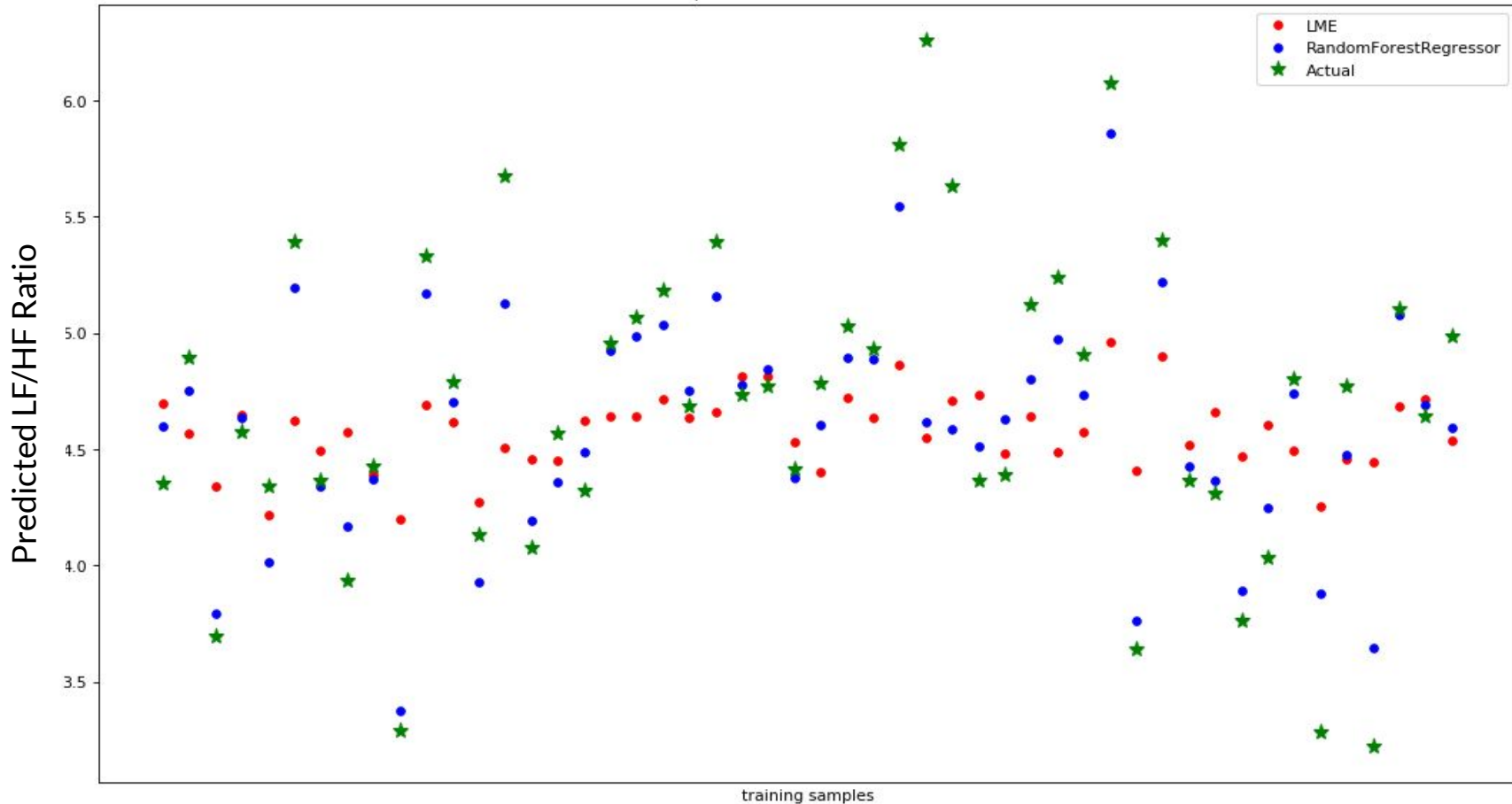
# Results



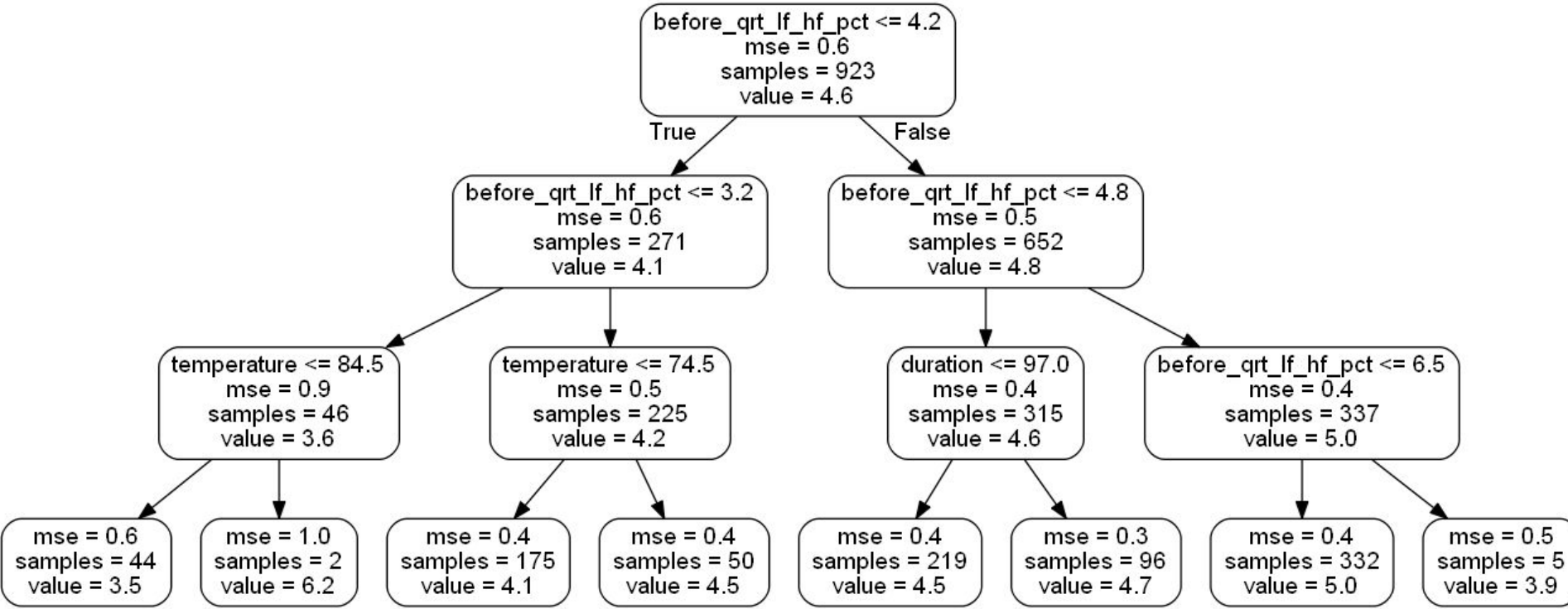
	LME	Before/After LME	Individual RF	Team RF	Before/After RF
QRT Index	164.14%	49.52%	529.88%	197.63%	34.97%
QRT %	245.59%	458.00%	11690.59%	238.23%	260.40%
LF	838.31%	560.00%	150.96%	4766.27%	29.21%
HF	139.66%	120.00%	750.30%	218.05%	78.37%
Lf/Hf %	115.21%	78.50%	258.06%	196.44%	11.55%
VLF	126.00%	99.00%	560.04%	888.17%	206.93
RMSSD %	493.54%	268.00%	35372.41%	3264.01%	10.59
RMSSD	120.77%	99.00%	122.32%	506.32%	30.01%

Mean Absolute Percentage Error for all Models and HRV attributes


Comparison of Model Predictions



The Random Forest predicts much closer to the actual values compared to the Mixed Effects Model



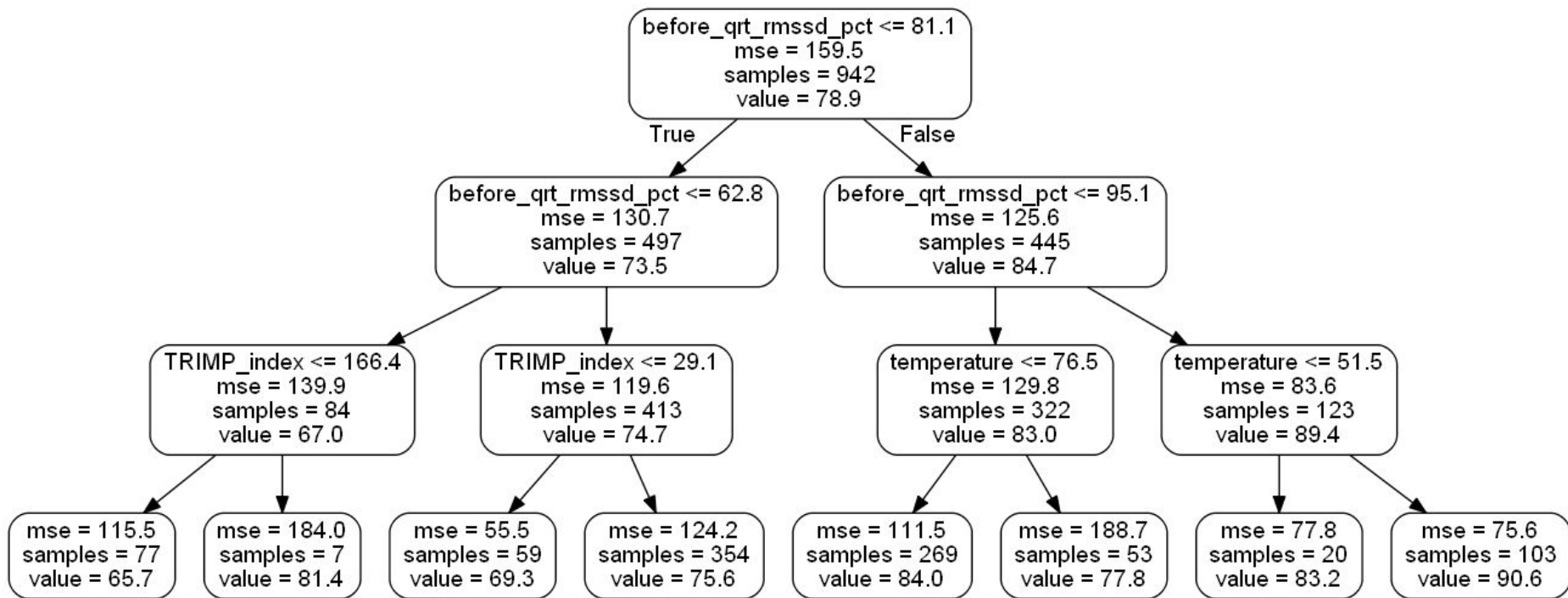
A single tree from the random forest model predicting LF/HF Ratio after an Activity



	LME	Before/After LME	Individual RF	Team RF	Before/After RF
QRT Index	164.14%	49.52%	529.88%	197.63%	34.97%
QRT %	245.59%	458.00%	11690.59%	238.23%	260.40%
LF	838.31%	560.00%	150.96%	4766.27%	29.21%
HF	139.66%	120.00%	750.30%	218.05%	78.37%
Lf/Hf %	115.21%	78.50%	258.06%	196.44%	11.55%
VLF	126.00%	99.00%	560.04%	888.17%	206.93
RMSSD %	493.54%	268.00%	35372.41%	3264.01%	10.59
RMSSD	120.77%	99.00%	122.32%	506.32%	30.01%

Mean Absolute Percentage Error for all Models and HRV attributes





A single tree from the random forest model predicting RMSSD % after an Activity



# Results

- Linear Mixed Effects Model does not perform well
- Random Forest
  - does not perform well on individual player data or team data
  - performs better on data using former HRV measurements as a predictor



# Conclusions

- Biometrics often do not perform well under linear analysis
- More individual player data is needed to perform predictions on individual data
  - If every player increased their usage of the Firstbeat tracker, this should be revisited
- It makes sense to use the pre-activity HRV measurements as this is a baseline from which the player is potentially moving from.
- This data shows that when in certain ranges for each HRV, temperature has a varying effect on post-activity HRV.
- A player in a well recovered state, subjected to high temperatures will be in a less recovered state post-activity relative if they had performed that activity at moderate temperatures
  -



# Continued Work

- Neural Network
- Improve data quantity and quality