Modeling and Prediction of Athletic Readiness based on Sleep and Recovery Patterns

Krina Khakhariya, Urjit Mehta, Brijesh Munjiyasara, Khushi Agrawal, and Jafri Syed Mujtaba

School of Engineering and Applied Science, Ahmedabad University

Email: {krina.k, urjit.m, brijesh.m, khushi.a2, and jafri.h}@ahduni.edu.in

Team: **Panchtron**

Ahmedabad University

April 15, 2025

Problem Statement

Ahmedabad University

Predicting Athletic Readiness Using Sleep & Recovery Data

In collegiate basketball, athletes experience fatigue due to frequent games, travel, training, academics, and social commitments. Fatigue negatively affects sleep patterns and recovery, which in turn affects your athletic performance. The goal of this project is to analyze sleep and recovery data and predict RSImod (Readiness Measure), a key indicator of an athlete's readiness for competition.

- Frequent games, travel, and academic stress contribute to poor sleep and slower recovery, affecting athletic performance
- Sleep patterns and recovery metrics are examined to identify trends influencing readiness
- Insights from the data help optimize training and rest schedules for improved competition readiness

Instructor's Feedback

Feedback Provided:
Increase the dataset to enhance model performance and data constraints.

What We Did:

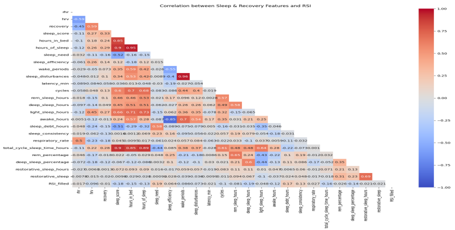
- Preserved feature associations and natural distributions
- Applied **CTGAN** (Conditional Tabular GAN) to create artificial data
- Managed missing data and disparate data types (categorical + numeric)

Effect:
Allowed for further ML tests, enhanced dataset completeness, and facilitated privacy. Synthetic data, however, did not considerably increase model accuracy — ongoing testing.

Data Imputation

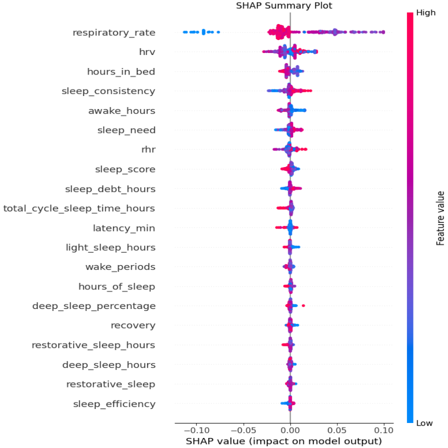
- Dataset had 90% rows empty in Target variable RSI.
- RSI was backward filled using a 13 day window.
- Other Numerical variables were imputed using MICE.

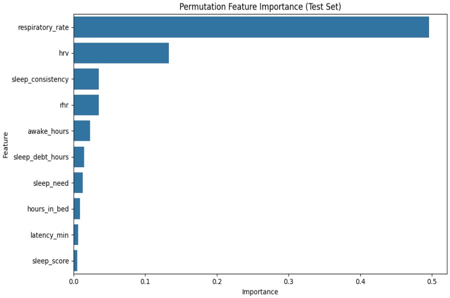
Feature	Original Mean	Imputed Mean	% Mean Error	Original Std	Imputed Std	% Std Error
rhr	59.69	59.59	0.16	9	8.27	8.17
hrv	84.07	85.8	2.07	36.11	33.63	6.86
recovery	59.55	61.45	3.19	22.66	19.06	15.91
sleep_score	76.39	76.27	0.17	18.55	15.65	15.67
hours_in_bed	7.78	7.76	0.33	1.93	1.62	15.91
hours_of_sleep	6.89	6.87	0.26	1.65	1.38	16.28
sleep_need	8.91	8.92	0.09	1.12	0.97	13.2
sleep_efficiency	88.87	88.96	0.1	6.3	5.32	15.48
wake_periods	14.87	14.19	4.56	6.61	5.92	10.42
sleep_disturbances	11.63	11.7	0.6	5.61	5.03	10.29
latency_min	2.89	2.76	4.21	5.28	4.2	20.43
cycles	5.3	5.24	1.26	2.18	1.83	16.17
rem_sleep_hours	2.04	2.02	1.2	0.79	0.67	14.43
deep_sleep_hours	1.37	1.36	0.74	0.46	0.39	13.9
light_sleep_hours	3.47	3.49	0.48	1.21	1.04	14.35
awake_hours	0.9	0.89	0.7	0.62	0.54	12.8
sleep_debt_hours	0.98	0.98	0.64	0.73	0.62	15.83
sleep_consistency	62.59	60.69	3.04	14.97	13.45	10.13
respiratory_rate	16.46	16.51	0.36	1.58	1.63	3.14
total_cycle_sleep_time_hours	7.18	7.15	0.38	1.79	1.51	15.63
rem_percentage	26.37	26.12	0.96	8.5	7.25	14.69
deep_sleep_percentage	17.79	17.75	0.18	4.79	4.03	15.9
restorative_sleep_hours	3.41	3.42	0.26	1.11	0.92	17.57
restorative_sleep	44.14	44.34	0.45	11.07	9.19	17
RSI	0.38	0.37	1.68	0.08	0.08	1.93

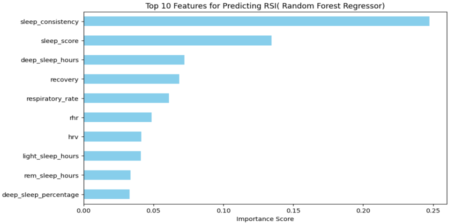


- Comparison table shows, % error in mean and sd, of original vs imputed data.
- No one variable has high corr with RSI.

Feature Importance







Methodology

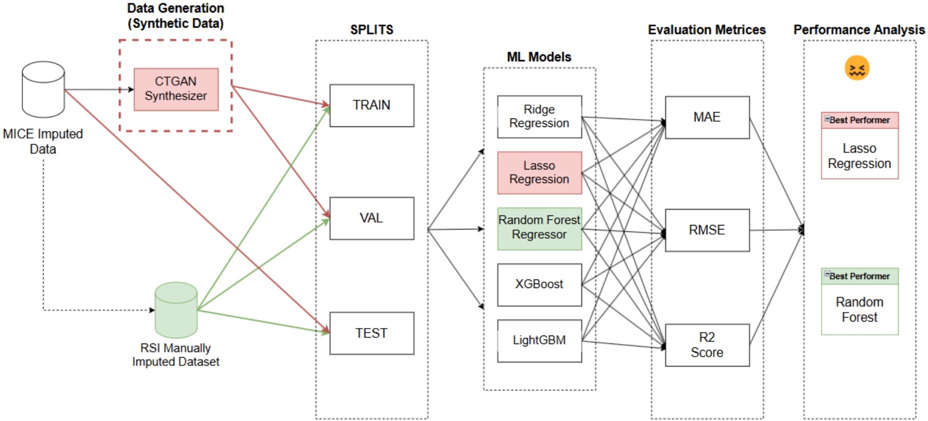


Figure 1: Different Approaches' Pipeline

Results

Ahmedabad University

Table 1: Imputed_MICE_dataset with RSI

Model	MAE	RMSE	R2 Score
Ridge	0.00278	0.000159	0.898314
Lasso	0.019029	0.000919	0.411488
Random Forest	0.010224	0.000392	0.74909
XGBoost	0.010022	0.000373	0.76089
LightGBM	0.008216	0.000307	0.80346

Table 2: Imputed_MICE_dataset wo RSI m

Model	MAE	RMSE	R2 Score
Ridge	0.055	0.005	0.1051
Lasso	0.057	0.0054	0.0319
Random Forest	0.0442	0.0033	0.3974
XGBoost	0.0456	0.0036	0.3488
LightGBM	0.0447	0.0036	0.3555

Results (continue)

Table 3: Model Performance on imputed_MICE dataset while synthetic test data (600 Rows)

Model File	MAE	RMSE	R2 Score
imputed_MICE_Ridge.pkl	1.491281	2.597637	-1547.213526
imputed_MICE_Lasso.pkl	0.025846	0.001678	-0.000043
imputed_MICE_RandomForest.pkl	0.031116	0.001825	-0.08744
imputed_MICE_XGBoost.pkl	0.029995	0.001883	-0.122393
imputed_MICE_LightGBM.pkl	0.028352	0.001728	-0.029748

Table 4: Model Performance on Synthetic Dataset(10000 Rows) while imputed MICE test data

Model File	MAE	RMSE	R2 Score
Synthetic_Dataset_Ridge.pkl	0.050864	0.003792	-1.259779
Synthetic_Dataset_Lasso.pkl	0.031938	0.001896	-0.130108
Synthetic_Dataset_RandomForest.pkl	0.081423	0.008006	-3.771541
Synthetic_Dataset_XGBoost.pkl	0.089097	0.009784	-4.8316
Synthetic_Dataset_LightGBM.pkl	0.040605	0.00274	-0.63335

Conclusion & Future Scope

Ahmedabad University

Conclusion

- Performed systematic tuning and evaluation to enhance the performance and reliability of all models, ensuring better accuracy and generalizability across different datasets; despite this, the models are not giving satisfactory outcomes

Future Scope

- Optimize models using hyperparameter tuning and ensemble approaches for better performance and stability.
- Investigate more sophisticated imputation and data augmentation methods to improve generalizability.

References

1. Senbel, S., Sharma, S., Raval, M. S., Taber, C., Nolan, J., Artan, N. S., Ezzeddine, D., & Kaya, T. (2022). *Impact of sleep and training on game performance and injury in division-1 women's basketball amidst the pandemic*. IEEE Access, 10, 15516–15527.
2. Sargent, C., Lastella, M., Halson, S. L., & Roach, G. D. (2021). *How much sleep does an elite athlete need?* International Journal of Sports Physiology and Performance, 16(12), 1746–1757.
3. Mah, C. D., Mah, K. E., Kezirian, E. J., & Dement, W. C. (2011). *The effects of sleep extension on the athletic performance of collegiate basketball players*. Sleep, 34(7), 943–950.
4. Burke, T. M., Lisman, P. J., Maguire, K., Skeiky, L., Choynowski, J. J., Capaldi, V. F., Wilder, J. N., Brager, A. J., Dobrosielski, D. A., et al. (2020). *Examination of sleep and injury among college football athletes*. The Journal of Strength & Conditioning Research, 34(3), 609–616.
5. Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). *Modeling Tabular Data Using Conditional GAN*. Advances in Neural Information Processing Systems (NeurIPS), 32, 7335–7345.