**Predicting NDVI which is the vegetation and greenness on area on the beach using machine learning methods (LSTM, SVR, RF, MLR) and regression analysis.**

**Submitted by**

**Submitted to**

**Dated**

**Predicting NDVI which is the vegetation and greenness on area on the beach using machine learning methods (LSTM, SVR, RF, MLR) and regression analysis.**

Contents

[Results 1](#_Toc149880812)

[Discussion 31](#_Toc149880813)

[Conclusion 34](#_Toc149880814)

# Results

**THIS TABLE CONTAINS ALL RESULTS NEEDED**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Analysis Type** | **Factors** | **Time-Lag** | **Data** | **Area** | **R Square** | **Significance F** |
| Regression | ALL | No | NDVI | 2 | 0.46829 | 5.42E-07 |
| Regression | ALL | No | NDVI | 4 | 0.386758 | 3.34E-05 |
| Regression | ALL | No | NDVI | 5 | 0.322276 | 0.000541 |
| Regression | ALL | No | NDVI | 6 | 0.172342 | 0.083414 |
| Regression | ALL | 1 | NDVI | 2 | 0.572036 | 1.23E-09 |
| Regression | ALL | 1 | NDVI | 4 | 0.383863 | 4.7E-05 |
| Regression | ALL | 1 | NDVI | 5 | 0.303183 | 0.001327 |
| Regression | ALL | 1 | NDVI | 6 | 0.216045 | 0.025823 |
| Regression | Main | Yes | NDVI | 2 | 0.548729 | 1.31E-11 |
| Regression | ALL | Yes | NDVI-Diff | 2 | 0.235246 | 0.014203 |
| Regression | ALL | Yes | NDVI-Diff | 4 | 0.089356 | 0.525059 |
| Regression | ALL | Yes | NDVI-Diff | 5 | 0.11163 | 0.355459 |
| Regression | ALL | Yes | NDVI-Diff | 6 | 0.140586 | 0.193474 |
| Regression | ALL | 2 | NDVI | 2 | 0.592709 | 4.25E-10 |
| Regression | ALL | 2 | NDVI | 4 | 0.385236 | 5.43E-05 |
| Regression | ALL | 2 | NDVI | 5 | 0.303302 | 0.001528 |
| Regression | ALL | 2 | NDVI | 6 | 0.263004 | 0.006364 |
| **Regression** | **ALL** | **2** | **Average** | **2** | **0.704709** | **3.01E-14** |
| **Regression** | **ALL** | **2** | **Average** | **4** | **0.512741** | **7.83E-08** |
| **Regression** | **ALL** | **2** | **Average** | **5** | **0.403524** | **2.37E-05** |
| **Regression** | **ALL** | **2** | **Average** | **6** | **0.344057** | **0.000313** |
| Regression | ALL | 3 | NDVI | 2 | 0.481241 | 6.3E-07 |
| Regression | ALL | 3 | NDVI | 4 | 0.307803 | 0.001498 |
| Regression | ALL | 3 | NDVI | 5 | 0.233382 | 0.018365 |
| Regression | ALL | 3 | NDVI | 6 | 0.189315 | 0.064868 |
| Regression | ALL | 3 | Average | 2 | 0.591268 | 7.11E-10 |
| Regression | ALL | 3 | Average | 4 | 0.406473 | 2.59E-05 |
| Regression | ALL | 3 | Average | 5 | 0.310467 | 0.001358 |
| Regression | ALL | 3 | Average | 6 | 0.292383 | 0.002616 |

In this detailed examination, we dig into the relationships between the Normalized Difference Vegetation Index (NDVI), multiple influencing factors, time lags, data areas, and averaging techniques. The basic R-squared values and significance levels (F-statistics) are carefully inspected to measure the strength and significance of these relationships. Our underlying request focuses on the role of time lag in forming NDVI, enveloping all elements (Regression: ALL). The results uncover what differing delays mean for the R-squared values and the levels of significance:

* At Time Lag 2: R-squared = 0.46829, Significance F = 5.42E-07
* At Time Lag 4: R-squared = 0.386758, Significance F = 3.34E-05
* At Time Lag 5: R-squared = 0.322276, Significance F = 0.000541
* At Time Lag 6: R-squared = 0.172342, Significance F = 0.083414

The presentation of a time lag of 1 (Time-Lag: 1) yields the accompanying outcomes:

* At Time Lag 2: R-squared = 0.572036, Significance F = 1.23E-09
* At Time Lag 4: R-squared = 0.383863, Significance F = 4.7E-05
* At Time Lag 5: R-squared = 0.303183, Significance F = 0.001327
* At Time Lag 6: R-squared = 0.216045, Significance F = 0.025823

These discoveries give an exhaustive perspective on what the determination of delays means for the connection among NDVI and other impacting factors.

The following spotlight is on the impact of various information regions with regards to NDVI regression analysis. Under the consideration of all elements and the usage of NDVI without indicating an information region (Data: NDVI), we observed the accompanying:

Without indicating a specific information region (No Data Area): R-squared = 0.548729, Significance F = 1.31E-11

Utilizing NDVI-Diff as the information region: R-squared = 0.235246, Significance F = 0.014203

This correlation highlights the significant effect that the selection of information region can have on the connection among NDVI and the different impacting factors.

Finally, our examination goes to the impact of various averaging strategies on NDVI relapse investigation. By taking into account all elements (Regression: ALL) and utilizing a scope of averaging techniques, we revealed the accompanying results:

* Using Averaging Method 2: R-squared = 0.704709, Significance F = 3.01E-14
* Using Averaging Method 4: R-squared = 0.512741, Significance F = 7.83E-08
* Using Averaging Method 5: R-squared = 0.403524, Significance F = 2.37E-05
* Using Averaging Method 6: R-squared = 0.344057, Significance F = 0.000313

These outcomes underscore the significant effect that the decision of averaging techniques can apply on the connection among NDVI and the different impacting factors with regards to regression analysis.

**PLEASE NOTE THAT**: THERE ARE MORE RESULTS AND TABLES IN EXCEL FILES

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | No | NDVI | 2 | 0.46829 | 5.42E-07 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.684317 |
| R Square | 0.46829 |
| Adjusted R Square | 0.410134 |
| Standard Error | 0.047675 |
| Observations | 72 |

In this analysis, we find that our model has a genuinely strong association with the dependent variable, set apart by a multiple R-value of 0.684317. The R-squared worth of 0.46829 implies that our model records for around 46.83% of the differences we see in the dependent variable, showing a significant capacity to make sense of the information. The changed R-squared value, 0.410134, affirms that in any event, when we consider the quantity of indicators utilized in our model, it stays viable in making sense of the difference we see in the information. The standard error, measured at 0.047675, signifies the typical margin of error between our model's predictions and the actual data. This analysis is based on a dataset comprising 72 observations.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -9.28733 | 3.078414 | -3.01692 | 0.003661 |
| Temp(c) | 0.006294 | 0.006586 | 0.955599 | 0.342871 |
| WindSpeed(m/s) | -0.02384 | 0.015625 | -1.5259 | 0.131961 |
| Humidity(%) | -0.00249 | 0.002098 | -1.18728 | 0.239507 |
| Pressure(mbar) | 0.009438 | 0.002869 | 3.289834 | 0.001633 |
| CloudType(Ref) | -0.00175 | 0.011127 | -0.15762 | 0.875253 |
| WindDirection(Degrees) | 0.000174 | 0.00033 | 0.526835 | 0.600129 |
| Precipitation(cm) | 0.013495 | 0.022837 | 0.590934 | 0.556646 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | No | NDVI | 4 | 0.386758 | 3.34E-05 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.621899 |
| R Square | 0.386758 |
| Adjusted R Square | 0.319685 |
| Standard Error | 0.075526 |
| Observations | 72 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -15.2908 | 4.87678 | -3.13544 | 0.002591 |
| Temp(c) | 0.028201 | 0.010434 | 2.702887 | 0.008792 |
| WindSpeed(m/s) | -0.05795 | 0.024753 | -2.34097 | 0.022361 |
| Humidity(%) | 0.002528 | 0.003323 | 0.760721 | 0.449617 |
| Pressure(mbar) | 0.014854 | 0.004545 | 3.268351 | 0.001742 |
| CloudType(Ref) | 0.004052 | 0.017628 | 0.229887 | 0.818913 |
| WindDirection(Degrees) | 0.000411 | 0.000523 | 0.786021 | 0.434756 |
| Precipitation(cm) | -0.08204 | 0.036178 | -2.26772 | 0.02673 |

The analysis of these coefficients gives us valuable insights into how different factors affect the dependent variable. We find that the intercept, at -15.2908, has a significant impact on the dependent variable, backed by a low p-value of 0.002591. Among the predictors, "Temperature" and "Pressure" shine as they make notable contributions to the model with low p-values of 0.008792 and 0.001742, respectively. The t-stats of 2.702887 and 3.268351 for these variables show that they have a strong influence on the dependent variable. Then again, indicators like "Humidity," "Cloud Type," "Wind Direction," and "Precipitation" don't show genuinely huge connections since their p-values outperform the normal importance edge of 0.05. These discoveries assist us with understanding which factors assume a vital part in our model and which ones might have a limited effect.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | No | NDVI | 5 | 0.322276 | 0.000541 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.567693 |
| R Square | 0.322276 |
| Adjusted R Square | 0.24815 |
| Standard Error | 0.094954 |
| Observations | 72 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -16.1392 | 6.131242 | -2.63229 | 0.010618 |
| Temp(c) | 0.030179 | 0.013118 | 2.300621 | 0.024682 |
| WindSpeed(m/s) | -0.06083 | 0.031121 | -1.95477 | 0.054981 |
| Humidity(%) | 0.000433 | 0.004178 | 0.103542 | 0.917857 |
| Pressure(mbar) | 0.015851 | 0.005714 | 2.774135 | 0.007245 |
| CloudType(Ref) | 0.011946 | 0.022162 | 0.53903 | 0.591736 |
| WindDirection(Degrees) | 0.000403 | 0.000658 | 0.612425 | 0.542427 |
| Precipitation(cm) | -0.07802 | 0.045484 | -1.71541 | 0.091108 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | No | NDVI | 6 | 0.172342 | 0.083414 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.415141 |
| R Square | 0.172342 |
| Adjusted R Square | 0.081817 |
| Standard Error | 0.059892 |
| Observations | 72 |

The analysis in Table 1, where we considered all available factors without any time lag and focused on the Normalized Difference Vegetation Index (NDVI) in a specific context (Area 6), demonstrates that our model can explain approximately 17.23% of the variations in NDVI. While this level of explanation is moderate, the Significance F value, which helps assess the model's overall significance, exceeds the typical threshold of 0.05. In Table 2, we find additional insights. The Multiple R value of 0.415141 indicates a moderate correlation between the variables, and the R-squared value reaffirms the model's ability to explain the data. However, the lower Adjusted R Square, which considers the number of predictors, hints at potential limitations in our model's robustness. The Standard Error of 0.059892 reflects the typical gap between predicted and actual values. Our analysis is based on a dataset with 72 observations, suggesting that further investigation or adjustments may be needed to enhance the model's explanatory power and statistical significance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -5.56922 | 3.867282 | -1.44009 | 0.154717 |
| Temp(c) | 0.010396 | 0.008274 | 1.25649 | 0.213505 |
| WindSpeed(m/s) | -0.03301 | 0.019629 | -1.68164 | 0.097513 |
| Humidity(%) | -0.00033 | 0.002635 | -0.12594 | 0.900173 |
| Pressure(mbar) | 0.005605 | 0.003604 | 1.555235 | 0.124822 |
| CloudType(Ref) | 0.00686 | 0.013979 | 0.490713 | 0.625308 |
| WindDirection(Degrees) | 0.000371 | 0.000415 | 0.894827 | 0.374233 |
| Precipitation(cm) | -0.02205 | 0.028689 | -0.76848 | 0.445027 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI | 2 | 0.572036 | 1.23E-09 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.75633 |
| R Square | 0.572036 |
| Adjusted R Square | 0.524484 |
| Standard Error | 0.042534 |
| Observations | 71 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -7.76763 | 2.752215 | -2.82232 | 0.006372 |
| X Variable 1 | 0.011286 | 0.005952 | 1.896371 | 0.062498 |
| X Variable 2 | -0.0333 | 0.014044 | -2.37093 | 0.020815 |
| X Variable 3 | -0.00251 | 0.001991 | -1.26175 | 0.211691 |
| X Variable 4 | 0.007847 | 0.002562 | 3.062398 | 0.003227 |
| X Variable 5 | 0.010648 | 0.009972 | 1.067842 | 0.289667 |
| X Variable 6 | 0.000264 | 0.000301 | 0.876769 | 0.383943 |
| X Variable 7 | 0.006926 | 0.020872 | 0.331837 | 0.741114 |

The coefficients we've obtained from our regression analysis provide us with some valuable insights into how different variables interact. We find that the intercept, which is at -7.76763, plays a significant role in affecting the dependent variable. This is supported by a low p-value of 0.006372, indicating its importance. Particularly, "X Variable 4" stands out as a powerful predictor with a high t-stat of 3.062398 and a very low p-value of 0.003227, suggesting it has a strong impact on the dependent variable. "X Variable 2" also appears to be influential, with a t-stat of -2.37093 and a p-value of 0.020815. On the other hand, variables like "X Variable 1," "X Variable 3," "X Variable 5," "X Variable 6," and "X Variable 7" are not found to have the remarkable according to their specified values.Our analysis included all available factors and introduced a time lag (Time-Lag: Yes) while focusing on the Normalized Difference Vegetation Index (NDVI) in Area 2. The R-squared value of 0.572036 suggests that our model can explain a substantial portion of the variance in NDVI, highlighting its ability to make sense of the data. Additionally, the low Significance F value of 1.23E-09 underscores the statistical significance of our model, confirming its reliability in describing the relationships between the variables.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI | 4 | 0.383863 | 4.7E-05 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.619567 |
| R Square | 0.383863 |
| Adjusted R Square | 0.315403 |
| Standard Error | 0.075889 |
| Observations | 71 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -12.7345 | 4.910491 | -2.59333 | 0.011805 |
| X Variable 1 | 0.025579 | 0.010619 | 2.408817 | 0.018941 |
| X Variable 2 | -0.06289 | 0.025057 | -2.50986 | 0.01466 |
| X Variable 3 | 0.000701 | 0.003552 | 0.197278 | 0.844245 |
| X Variable 4 | 0.012547 | 0.004572 | 2.744575 | 0.007884 |
| X Variable 5 | 0.008954 | 0.017791 | 0.503265 | 0.616533 |
| X Variable 6 | 5.93E-05 | 0.000536 | 0.110508 | 0.912358 |
| X Variable 7 | -0.0655 | 0.03724 | -1.75893 | 0.083445 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI | 5 | 0.303183 | 0.001327 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.550621 |
| R Square | 0.303183 |
| Adjusted R Square | 0.225759 |
| Standard Error | 0.096327 |
| Observations | 71 |

The intercept, which is at -12.3791, hovers around the threshold of statistical significance (p = 0.05138), hinting at a potential influence on the dependent variable. Interestingly, "X Variable 2" emerges as a meaningful contributor with a t-stat of -2.03583 and a p-value of 0.045978, signifying its role in affecting the dependent variable. Additionally, "X Variable 4" stands out with a t-stat of 2.10286 and a p-value of 0.039479. "X Variable 1" also suggests some significance (p = 0.050784) with a t-stat of 1.991336. On the flip side, variables like "X Variable 3," "X Variable 5," "X Variable 6," and "X Variable 7" do not appear to have statistically significant connections, as their p-values surpass the typical threshold of 0.05.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -12.3791 | 6.232945 | -1.98607 | 0.05138 |
| X Variable 1 | 0.02684 | 0.013478 | 1.991336 | 0.050784 |
| X Variable 2 | -0.06475 | 0.031805 | -2.03583 | 0.045978 |
| X Variable 3 | 0.001257 | 0.004509 | 0.278794 | 0.781316 |
| X Variable 4 | 0.012202 | 0.005803 | 2.10286 | 0.039479 |
| X Variable 5 | 0.008664 | 0.022583 | 0.383647 | 0.702532 |
| X Variable 6 | -5.7E-05 | 0.000681 | -0.08345 | 0.933762 |
| X Variable 7 | -0.06462 | 0.047269 | -1.36712 | 0.176447 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI | 6 | 0.216045 | 0.025823 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.464807 |
| R Square | 0.216045 |
| Adjusted R Square | 0.128939 |
| Standard Error | 0.058671 |
| Observations | 71 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -4.66439 | 3.796407 | -1.22863 | 0.223779 |
| X Variable 1 | 0.012146 | 0.00821 | 1.479537 | 0.14398 |
| X Variable 2 | -0.02596 | 0.019372 | -1.34014 | 0.185012 |
| X Variable 3 | -0.00244 | 0.002746 | -0.88839 | 0.377708 |
| X Variable 4 | 0.004826 | 0.003534 | 1.365335 | 0.177003 |
| X Variable 5 | 0.01592 | 0.013755 | 1.15741 | 0.251474 |
| X Variable 6 | 2.38E-05 | 0.000415 | 0.057369 | 0.954432 |
| X Variable 7 | -0.02004 | 0.028791 | -0.69598 | 0.488999 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | Main Factors | 2 | 0.548729 | 1.31E-11 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.740762 |
| R Square | 0.548729 |
| Adjusted R Square | 0.528523 |
| Standard Error | 0.042353 |
| Observations | 71 |

In the world of regression analysis, these statistics hold some fascinating revelations. Our Multiple R score, which sits at 0.740762, tells us that there's a robust connection between our independent and dependent variables. As we delve deeper, the R Square value of 0.548729 unveils that our model has the ability to clarify roughly 54.87% of the variations in the dependent variable. This is a substantial level of explanatory power. Looking at the Adjusted R Square, which is slightly lower at 0.528523, we see that it retains its effectiveness even when we consider the number of predictors in our model. This suggests that our model is resilient and adaptable. The Standard Error, pegged at 0.042353, gives us an idea of how accurate our predictions tend to be. It indicates that our model is pretty good at making predictions with a fair degree of precision. With a dataset comprising 71 observations, our analysis is firmly grounded in a robust pool of data, reinforcing the reliability of our findings.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | -7.91345 | 2.117679 | -3.73685 | 0.000388 |
| X Variable 1 | 0.011161 | 0.001661 | 6.718905 | 4.81E-09 |
| X Variable 2 | -0.03384 | 0.013502 | -2.50608 | 0.014642 |
| X Variable 3 | 0.007883 | 0.00203 | 3.882793 | 0.000239 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI-Diff | 2 | 0.235246 | 0.014203 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.485021 |
| R Square | 0.235246 |
| Adjusted R Square | 0.150273 |
| Standard Error | 0.031076 |
| Observations | 71 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | 1.320999 | 2.010809 | 0.656949 | 0.513607 |
| X Variable 1 | 0.006047 | 0.004348 | 1.3907 | 0.169209 |
| X Variable 2 | -0.01135 | 0.010261 | -1.10646 | 0.272736 |
| X Variable 3 | -0.00146 | 0.001872 | -0.78043 | 0.438059 |
| X Variable 4 | 0.000736 | 0.001455 | 0.506233 | 0.61446 |
| X Variable 5 | 0.011357 | 0.007285 | 1.558878 | 0.124035 |
| X Variable 6 | -0.01162 | 0.015249 | -0.76226 | 0.448749 |
| X Variable 7 | 0.000155 | 0.00022 | 0.705858 | 0.482878 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI-Diff | 4 | 0.089356 | 0.525059 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.298924 |
| R Square | 0.089356 |
| Adjusted R Square | -0.01183 |
| Standard Error | 0.041185 |
| Observations | 71 |

We delved into a thorough examination, taking into account all possible factors and introducing a time lag. With a dataset of 71 observations, our examination has serious areas of strength for a, reaffirming the unwavering quality of our discoveries and adding an exceptional viewpoint to our investigation. Our focus was on NDVI-Diff within the captivating realm of Area 4. The outcome revealed that our model, with an R-squared value of 0.089356, explains about 8.94% of the variations in the dependent variable. This suggests a rather modest level of explanatory power. Interestingly, the Significance F value, which defies the usual significance threshold at 0.525059, raises some questions about the overall statistical significance of our model. Taking a closer look at Table 2, we find a moderate correlation (Multiple R), but the Adjusted R Square falls notably below the R Square. This indicates that when we account for the number of predictors, the model's effectiveness diminishes, suggesting some potential weaknesses. The Standard Error of 0.041185 provides an understanding of how closely our predictions align with actual values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | 2.212925 | 2.664938 | 0.830385 | 0.409457 |
| X Variable 1 | -0.0008 | 0.005763 | -0.13889 | 0.889982 |
| X Variable 2 | -0.00822 | 0.013598 | -0.60475 | 0.547517 |
| X Variable 3 | -0.00208 | 0.002481 | -0.839 | 0.404642 |
| X Variable 4 | -0.00052 | 0.001928 | -0.26829 | 0.78935 |
| X Variable 5 | 0.003096 | 0.009655 | 0.320603 | 0.749573 |
| X Variable 6 | 0.007804 | 0.02021 | 0.38613 | 0.700701 |
| X Variable 7 | -0.00024 | 0.000291 | -0.82082 | 0.414845 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI-Diff | 5 | 0.11163 | 0.355459 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.33411 |
| R Square | 0.11163 |
| Adjusted R Square | 0.012922 |
| Standard Error | 0.050474 |
| Observations | 71 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | 3.723449 | 3.266009 | 1.140061 | 0.258577 |
| X Variable 1 | -0.00314 | 0.007063 | -0.4452 | 0.657699 |
| X Variable 2 | -0.00427 | 0.016666 | -0.25596 | 0.798813 |
| X Variable 3 | -0.00363 | 0.003041 | -1.19224 | 0.237637 |
| X Variable 4 | 0.000964 | 0.002363 | 0.408116 | 0.684572 |
| X Variable 5 | -0.00348 | 0.011833 | -0.29367 | 0.769972 |
| X Variable 6 | 0.01247 | 0.024768 | 0.503449 | 0.616405 |
| X Variable 7 | -0.00045 | 0.000357 | -1.25526 | 0.214021 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis Type | Factors | Time-Lag | Data | Area | R Square | Significance F |
| Regression | ALL | Yes | NDVI-Diff | 6 | 0.140586 | 0.193474 |

|  |  |
| --- | --- |
| Regression Statistics | |
| Multiple R | 0.374948 |
| R Square | 0.140586 |
| Adjusted R Square | 0.045096 |
| Standard Error | 0.037359 |
| Observations | 71 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | P-value |
| Intercept | 0.85176 | 2.417394 | 0.352347 | 0.725754 |
| X Variable 1 | 0.002032 | 0.005228 | 0.388682 | 0.698822 |
| X Variable 2 | 0.006541 | 0.012335 | 0.53027 | 0.597789 |
| X Variable 3 | -0.00074 | 0.002251 | -0.33093 | 0.741799 |
| X Variable 4 | -0.00191 | 0.001749 | -1.08963 | 0.280028 |
| X Variable 5 | 0.008781 | 0.008758 | 1.002593 | 0.319893 |
| X Variable 6 | 0.000659 | 0.018333 | 0.035936 | 0.971447 |
| X Variable 7 | -0.00033 | 0.000264 | -1.25044 | 0.215762 |

**THIS TABLE CONTAINS ALL NEEDED RESULTS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Area** | **Method** | **MAE** | **MSE** | **RMSE** | **R-Squared** |
| MA 2 | LSTM EPOCHS 200 | 0.122696 | 0.025648 | 0.160149 | 0.592917 |
| MA 2 | OLS | 0.037284 | 0.002134 | 0.046199 | 0.508 |
| MA 2 | MLR | 0.036148 |  | 0.046286 | 0.428837 |
| MA 2 | RF | 0.022931 |  | 0.034954 | 0.674266 |
| MA 2 | SVR | 0.029911 |  | 0.042140 | 0.526585 |
| MA 4 | LSTM EPOCHS 200 | 0.147490 | 0.036470 | 0.190973 | 0.500608 |
| MA 4 | OLS | 0.062498 | 0.005365 | 0.073247 | 0.434 |
| MA 4 | MLR | 0.058816 |  | 0.073090 | 0.355897 |
| MA 4 | RF | 0.039835 |  | 0.059076 | 0.579215 |
| MA 4 | SVR | 0.046360 |  | 0.067099 | 0.457155 |
| MA 5 | LSTM EPOCHS 200 | 0.152525 | 0.041470 | 0.203642 | 0.463425 |
| MA 5 | OLS | 0.079744 | 0.008690 | 0.093221 | 0.365 |
| MA 5 | MLR | 0.073653 |  | 0.092339 | 0.278373 |
| MA 5 | RF | 0.048511 |  | 0.070457 | 0.579858 |
| MA 5 | SVR | 0.057844 |  | 0.080437 | 0.452410 |
| MA 6 | LSTM EPOCHS 200 | 0.158299 | 0.040177 | 0.200441 | 0.237097 |
| MA 6 | OLS | 0.046529 | 0.003072 | 0.055423 | 0.237 |
| MA 6 | MLR | 0.046291 |  | 0.057552 | 0.149896 |
| MA 6 | RF | 0.027611 |  | 0.041654 | 0.554671 |
| MA 6 | SVR | 0.035667 |  | 0.051998 | 0.306057 |
| MA 2 | LSTM EPOCHS 500 | 0.016445 | 0.000600 | 0.024502 | 0.990471 |
| MA 4 | LSTM EPOCHS 500 | 0.015301 | 0.000417 | 0.020411 | 0.9942955 |
| MA 5 | LSTM EPOCHS 500 | 0.039613 | 0.004992 | 0.070656 | 0.935406 |
| MA 6 | LSTM EPOCHS 500 | 0.047543 | 0.006047 | 0.077760 | 0.885182 |

In this comprehensive evaluation, we explored different methods for predicting outcomes in various areas (MA 2, MA 4, MA 5, MA 6) and assessed their performance using various metrics.

LSTM EPOCHS 200: In the context of MA 2, the LSTM model trained for 200 epochs demonstrated fairly accurate predictions. OLS (Ordinary Least Squares): In MA 2, the OLS method produced a lower MAE of 0.037284, signifying relatively precise predictions. However, it had a slightly higher RMSE of 0.046199. The R-squared value, standing at 0.508, suggests that it explains a significant portion of the variance. Comparable performance was found in other areas (MA 4, MA 5, MA 6). MLR (Multiple Linear Regression): MA 2's MLR resulted in a MAE of 0.036148 and an RMSE of 0.046286. The R-squared value of 0.428837 indicates moderate explanatory power. Similar performance was seen in MA 4 and MA 5.

RF (Random Forest): In MA 2, the Random Forest model achieved a lower MAE of 0.022931 and an RMSE of 0.034954, with a high R-squared value of 0.674266, signifying strong predictive capabilities. Good performance was also evident in other areas (MA 4, MA 5, MA 6).

LSTM EPOCHS 500: In MA 2 and other areas, the LSTM model with 500 training epochs stood out with remarkable performance. It achieved a very low MAE of 0.016445, an RMSE of 0.024502, and an exceptionally high R-squared value of 0.990471, indicating almost perfect predictive capabilities. These findings highlight the significance of choosing the right method and the number of training epochs for predictive accuracy. LSTM models at 500 epochs delivered outstanding results, while Random Forest models also demonstrated strong predictive capabilities. OLS and MLR offered moderate performance, and Support Vector Regression provided reasonable results. The choice of method should be tailored to the specific area and the desired level of predictive accuracy.

A graph of a red and blue line

Description automatically generated

A graph with red and blue lines

Description automatically generated

A graph of a line graph

Description automatically generated with medium confidence

A graph of a line

Description automatically generated with medium confidence

A graph of a graph

Description automatically generated with medium confidence

A graph of a line graph

Description automatically generated with medium confidence

A graph with red and blue lines

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

A graph with red and blue lines

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

A graph of a line graph

Description automatically generated with medium confidence

A graph with red and blue lines

Description automatically generated

A graph of a graph with red and blue lines

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

A graph of a graph

Description automatically generated with medium confidence

A graph with red and blue lines

Description automatically generated

A graph with red and blue lines

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

A graph with red and blue lines

Description automatically generated

A graph with red and blue lines

Description automatically generated

A graph showing a line of a graph

Description automatically generated with medium confidence

A graph showing a line of growth

Description automatically generated with medium confidence

A graph of a line graph

Description automatically generated with medium confidence

A graph showing a line of a graph

Description automatically generated with medium confidence

# Discussion

In this section, there is comprehensive information about the results given above. According to this, the results obtained from the analysis of predicting NDVI that is representing vegetation and greenness in a beach area through applying different learning methods and regression analysis are providing some valuable insights. Secondly, this analysis also involved different factors like data types, and time lags.

From the time-lag analysis, it can be observed that time-lags of 2,4,5 and 6 area were tested properly. through this, it is possible to assess their impact on NDVI prediction. From the time lag 2, the value of R-squared is about 0.46829, it indicates that the model explained about 46.83^ of the variances in NDVI. Furthermore, its significance F value is recorded at 5.42E-07 that suggesting that the results is significant statistically. When the value of time lag is increased, then R-square value tends to decrease with it. It indicates that shorter time lag is providing better predictions. Secondly, when the time lag is at 6, then its R-square value decreased to 0.172343 and its F-statistic is increases to 0.083414. Therefore, such values are suggesting that the model is less effective at this time lag.

From the analysis of Time-lag 1, it can be observed that its R-squared value was improved for shorter lags. It is suggesting that the previous time steps NDVI has put a huge impact on the current NDVI predictions. On the other hand, at time lag 2, the value of R-squared was increased to 0.572026 and it also contains a high significant F value of 1.23E-09. These values are indicated that by incorporating the previous time step can easily increase prediction performance.

According to the data type analysis, it can be observed that there is a distinction between NDVI-Diff and NDVI. Therefore, at time lag 2, when NDVI is used, then its R-square value was 0.548729 and its significance F value is extremely low with 1.31E-11. This value is showing that when NDVI us used as the data type, then it is extremely effective in predicting NDVI. On the other hand, when NDVI-Diff is used, then its R-square values were considerable low. Moreover, its significance F values are high too. Due to this, it suggests that when differences in NDVI values is used, then values might not be as effective for prediction.

This analysis is also expanded further by considering the combination of time lags and data types. Under these points, when it is present at time lag 2 with NDVI as the data type, then its R-squared value is reached to 0.704709 with significance F value of 3.01E-14. These values are showing that when combining a time lag of 2 with NDVI, the data type is producing highly effective model for NDVI predictions.

However, when averaging approach is used in the regression analysis, then promising results are produced. when it is present at time lag 2 with averaging approach, then its R-square value is recorded at 0.705709 and significance F is at 3.01E-14 that shows that the method outperforms other options. Through out the analysis, significance F values are indicating the overall significance of the models. When the significance F value is low, then it is suggesting more robust and significant models.

Secondly, the results obtained from regression analysis also provide valuable insights into the relationship between the various independent variables and dependent variables based on the context of predicting the NDVI in area 2. It can be observed that its intercept is present at -7.76763 and it is playing an important role in affecting the dependent variable NDVI. Furthermore, it us also supportive by a low p-value of 0.006372 indicating its importance. Secondly, this intercept also representing the baseline or initial value of NDVI when other independent variables are showing zero result. According to this case, when there is no other variable is present, then there is still a non-zero impact on NDVI is present.

Also, X-variable 4 stands out as an important predictor that contains a high t-stat value of 3.062398 and quite low value of p with 0.003227. Such results are indicating that x-variable 4 contains a huge impact on NDVI. Secondly, its high t-stat value is suggesting that its effect is somehow different from zero and the low p-value of the variable is confirming is statistical significance.

On the other hand, X variable 2, is also considered another influential variable with a t-state of -=2.37093 and its p-value is about 0.020815. It can be observed that this variable contains a negative t-stat value. Therefore, is significance level is indicating that it contains a significant impact on NDVI. Its negative sign is indicating that it contains inverse relationship with NDVI.

Furthermore, such variables like X variable 1, X variable 3, X variable 5, X variable 6, and X variable 7, are not considered as remarkable based on its specified values. It is suggesting that these variables are not affecting NDVI in area 2 that is already indicated with its t-stats and p-values.

However, overall model evaluation included all available factors and introduced a time lag. This model evaluation is indicating its ability to explain the variance level in NDVI. The reason behind it is that it shows R-square value of 0.572036 that suggests that our model has the ability to explain a substantial portion of the variance in NDVI. Under these facts, 57.20% of the variability in NDVI can be attributed to the variables present in the model. When R-square value is higher, then the model is indicating a better fit. Secondly, with low significance F value of 1.23E-09 will further underscores the statistical significance of the model. Therefore, this value is indicating that model as a whole is reliable for describing the main relationships between all variables. It also implies that the main relationship between the independent and dependent variables and NDVI are not random because they are provided by the model.

The main results of the analysis involves a huge examination of all different possible factors and introduced a time lag in the context of NDVI-Diff within Area 4 that provides valuable insights. The required dataset is consisted of about 71 observations and this analysis gained benefits from a reasonably sized dataset. It also adds robustness to the findings. On the other hand, the R-squared value is about 0.089356 that indicates that the model can easily explain approximately 8.94% of the variations in the independent variable NDVI-Diff. Secondly, it is also demonstrating some high level of explanatory powers because it is modest. It is suggesting that there are some different factors or variables that are not included in this model and contributing high to the variability in NDVI-Diff.

From the analysis of performance metrices of various methods used in the prediction of MA values present at various areas include MA 2, MA 4, MA5, and MA 6. Moreover, methods include LSTM, with 200 or 500 epochs, OLS, RF, and SVR. This metrics highlighted some important results.

From this, LSTM with 500 epochs is showing consistent results and outperforming other methods across all areas in terms of RMSE, MAE, MSE, and R-squared values. However, LSTM with 200 epochs also showed impressive results but still not much consistent. Furthermore, some traditional regression methods include MLR, and OLS contain high MSE, MAE and RMSE values and low R-squared values compared with LSTM. Due to this, it is suggesting that LSTM especially with 500 epochs is extremely effective for capturing the underlying patterns in the data.

Moreover, SVR, and RF methods are also performing well particularly in the areas MA 2 and 4. At these areas, the contains lower MAE, RMSE, and MSE values compared with MLR, and OLS. These methods are still too much behind from LSTM based on R-squared values. It can be noted that performance of variables varies across various area like in MA 2, LSTM with 500 epochs had achieved a huge accuracy rate with an R-square value of 0.990471. it indicates that the model is an excellent fit. However, in MA 6, LSTM with 200 epochs and other methods is showing quite low value of R-squared. Therefore, it suggests that predicting MA 4 in MA 6 is extremely challenging. The results are indicating that when number of epochs of LSTM is increased from 200 to 500, then the performance of LSTM is improved significantly. In various case, R-squared values approach or exceed 0.99. It indicates that the model is capturing a huge percentage of the variance in the data. The factor RMSE is involved in measuring the average magnitude of the errors present in the prediction of methods. From this, LSTM with 500 epochs is achieving the lowest RMSE value that indicates that it is providing highly accurate predictions compared with other methods. Additionally, the choice of prediction method must be depending on the specific area and context. Whenever, there is a need of high accuracy and precision, then LSTM with a sufficient number of epochs like 500 appears to be an ideal choice in the analysis. Moreover, some traditional regression methods like MLR and OLS performing worse compared with more advanced techniques in this context. Therefore, SVR and RF are highly competitive but not much accurate like LSTM when high epochs are used in it.

Now based on the required results obtained, it can be observed that LSTM machine learning method is standing out as the ideal and effective method for predicting NDVI in coastal areas like beaches. The results are highlighting that when LSTM contains 500 epochs, then it consistently outperformed the other machine learning methods because it is achieving high R-square values and low Root Mean Square Error values. These values are indicating that LSTM is highly reliable and efficient for capturing the complex temporal patterns present in NDVI data. Secondly, RF, SVR, and MLR are showing some promising results in certain cases. However, it contains some problems lagged behind LSTM with 500 epochs.

From the results, the Area MA 2 is considered as the most reliable and efficient area for NDVI predictions. At this area again LSTM with 500 epochs is demonstrating some exceptional result in performance with a huge value of R-squared about 0.990471. It indicates an excellent fit to the data. When the RMSE value is low, then it further underscore the accuracy of model while predicting NDVI changes in this specific area. Moreover, it is extremely important to note that the choice of area must aligned with the specific characteristics, and objectives of the study. From the results MA 2 is showing the best performance because the suitability of the area is depending on the research context and the availability of data.

# Conclusion

Summing up all the discussion from above, it is concluded that normalized difference vegetation index is considered a critical metric for monitoring greenness and vegetation in coastal areas like beaches. However, there is a need of providing accurate predictions of NDVI because it can provide valuable insights into the health and dynamics of coastal ecosystems. Moreover, it is also promoting benefits in environmental monitoring and land management. Under these points, this study had explored various machine learning methods include LSTM, SVR, RF, and MLR for predicting NDVI at different beach areas include MA 2, MA 4, MA 5, and MA 6.

From the key findings, include different factors. Therefore, the first one is time-lag analysis. This analysis is showing that when time lags are shorter, then it will lead towards better predictions of NDVI. The results had highlighted the importance of considering the impact of previous time steps applied on NDVI predictions. At time lag 1 hat is taken from previous time step showed improvement in predictive power of the model. It also suggests that the immediate past is playing an important role in NDVI changes. The result section had also analyzed data types by comparing NDVI and NDVI-difference. It can be observed that when NDVI is used as data types, then more effective results are obtained while predicting. Furthermore, with the combination of time lags of 2 with NDVI as the data type and this combination is producing highly effective models applied for NDVI predictions in coastal areas. Furthermore, by introducing an average approach in the regression analysis, the predictive performance is improving. Als with the combination of time lag of 2 with NDVI, the data type is providing highly accurate results with averaging providing with accurate predictions.

The regression analysis is focusing on the importance of various independent variables that include X variable 1 to X variable 7 for predicting NDVI. According to this, some variables like X variable 4 had provided some powerful prediction and contains less impact on the dependent variable. LSTM is an important deep learning method that consistently outperformed other machine learning methods for predicting NDVI include MLR, OLS, RF, and SVR in terms of predictive accuracy. Therefore, LSTM with 500 epochs had achieve remarkable high R-square values and low RMSEA. It indicates that its reliability in capturing complex temporal patterns is remarkable in NDVI data. On the other hand, the performance of the predictive models is also changing across various costal areas. Therefore, the Area MA 2 provided with the best performance for NDVI predictions. At this area, LSTM with 500 epochs provided with exceptionally accurate predictions with a huge R-square value that is quite reliable for this region.

The required findings of the research also contains some practical implications for using machine learning method in predicting NDVI in coastal areas. Whenever, there is a need to predict NDVI in coastal area, LSTM is ideal method for its when it contains a sufficient number of epochs like 500. It had showed some remarkable results and it is considered the most reliable method. This method also consistently outperforms other traditional regression methods and various machine learning techniques. Moreover, when NDVI is used as data types, then it is highly effective compared with other machine learning technique. Further when NDVI is used as the data type, then it is showing more accurate results and effective compared with NDVI-Diff for accurate predictions. However, implying absolute value of NDVI is extremely important for understanding vegetation changes in coastal regions. Through incorporating a time lag especially a lag 2 had improved prediction accuracy. It is suggesting that through considering the immediate past is extremely important for modelling NDVI changes effectively. Another point is that the choice of coastal area for NDVI prediction is also informed by the specific goals of the study and various characteristics of the area. Therefore, area MA 2 is ideal and well-suite for accurate NDVI predictions.

Beside these results, this research also contains some important limitations. From them, the first limitation is related to quality of data. The accuracy of NDVI prediction is mainly depending on the quality of the input data. If there is any inaccuracies or noise present in the data, then it can affect the reliability of the data. Moreover, the performance of machine learning models like LSTM is extremely sensitive to hyperparameter tuning. Therefore, such effectiveness of these models may vary according to the choice of made during the modelling process. Another limitation is that coastal areas are exhibiting some significant spatial and temporal variability in NDVI. Therefore, the required results for one specific area may not be directly applicable to other coastal regions. Due to this, these coastal areas may require different research. Secondly, the availability of historical and real-time NDVI data may vary from one area to another. This data will put a huge impact on the ability to implement these models effectively.

This study also opens the door for future research due to different points. The first point is related to model ensembles. Through investigating the potential benefits of model ensembles that will combine multiple models and strength can lead towards more accurate NDVI predictions. Furthermore, through leveraging additional data sources like meteorological data, and remote sensing data, it is possible to enhance the predictive power of the models. Secondly, by exploring advanced deep learning techniques, beyond LSTM like CNNs and hybrid models can enhance predictive accuracy and reliability. Lastly, extend the analysis by considering the spatial variability of NDVI in coastal area like beach can provide more comprehensive understanding of vegetation dynamics.