Diego Sarria

DSC Final Report

This project focuses on predicting surf conditions at the Pipeline surf spot using NOAA buoy data. The primary objective was to develop a model capable of translating offshore buoy readings into actionable surf forecasts for surfers. The combination of wave and meteorological data, alongside advanced machine learning techniques, formed the backbone of this effort. A PostgreSQL database was used to enable real-time data updates, ensuring predictions remain relevant and timely.

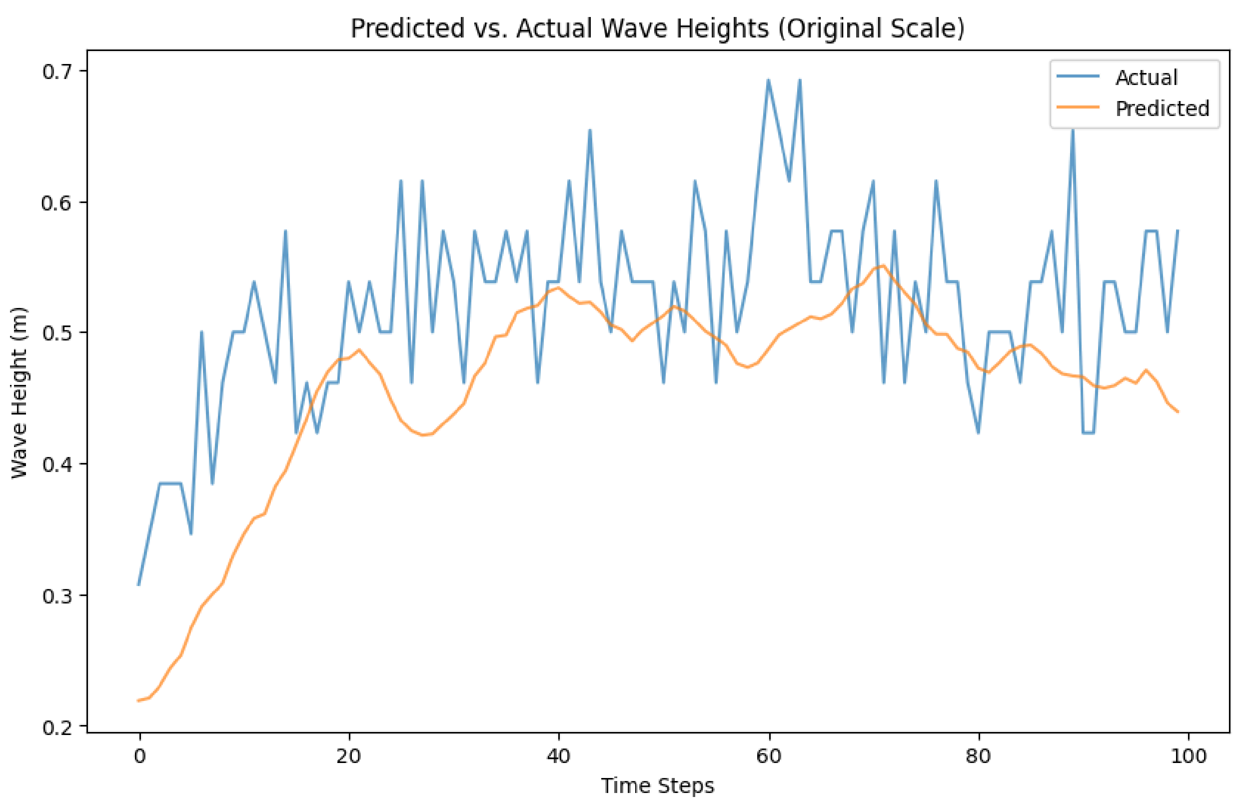
The data used in this project came from NOAA's buoy stations, including detailed wave metrics such as wave height, swell period, and direction, as well as meteorological data like wind speed, air temperature, and pressure. This dataset required extensive preprocessing, including handling missing values and aligning timestamps between wave and meteorological data. Exploratory analysis revealed significant correlations between wave height and period with surf conditions at Pipeline, justifying their use as primary features in the model.

One of the challenges of this project was accounting for the transformation of wave conditions from offshore buoy locations to the nearshore Pipeline surf spot. Physical transformations, such as wave transit time and shoaling, were applied to adjust the buoy observations to better reflect nearshore conditions. These adjustments were critical for ensuring the model's outputs aligned with the real-world behavior of waves as they travel from deep water to the shoreline.

An LSTM model was selected as the core predictive framework due to its ability to handle sequential data and learn long-term dependencies. The model was trained on preprocessed buoy data with a learning rate of α=0.006\alpha = 0.006, a batch size of 64, and 100 epochs. To ensure robust performance, early stopping and data splitting were employed to avoid overfitting. Additional hyperparameter tuning was conducted to optimize the model for this specific application.

The final model demonstrated strong performance, achieving a root mean squared error (RMSE) of 0.1305 on the test data. The model successfully predicted wave heights at Pipeline for future time intervals of 12 hours. These results validated the importance of data preprocessing and the effectiveness of using physical transformations to bridge the gap between offshore observations and nearshore conditions.

Despite these successes, there is room for improvement. The graph below compares the predicted wave heights to the actual wave heights for a subset of the test data. While the model captures some trends in wave height variation, there are noticeable discrepancies between the predictions and actual values, particularly for sharper peaks and troughs. These differences suggest that the model's ability to generalize across varying conditions is limited, likely due to the constraints of the training data.



To improve the model's reliability and better capture seasonal variations, it is essential to expand the dataset to cover a full year of observations. This is particularly important as wave behavior can change significantly across seasons, driven by factors like weather patterns and oceanic conditions. Additionally, future iterations could benefit from integrating data from multiple buoys and exploring advanced architectures, such as Temporal Fusion Transformers, to better capture complex temporal dependencies and improve overall performance. Developing a real-time dashboard using tools like Streamlit could also enhance usability, allowing stakeholders to visualize predictions alongside real-time data for informed decision-making.

Stakeholders for this project include surfers, coastal managers, and meteorological agencies. Accurate surf forecasts can improve safety and decision-making for these groups. However, the model must be used responsibly to avoid potential drawbacks, such as over-reliance on predictions without considering their uncertainties.

This project demonstrates the effective fusion of machine learning techniques with domain-specific physical knowledge to address a real-world challenge. The outcomes provide a strong foundation for future advancements in surf forecasting, with potential applications extending beyond surfing to other marine and coastal activities.

Citations

1. NOAA National Data Buoy Center, "Station 51101 - NDBC Standard Meteorological Buoy," [Online]. Available: https://www.ndbc.noaa.gov/station_page.php?station=51101.

2. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735. Available: https://www.researchgate.net/publication/13853244_Long_Short-term_Memory.