Predicting the Future of Kelp Forest Ecosystems in Southern California

Ishaan Chaturvedi

Abstract

Climate change has caused a continued increase in the frequency and intensity of heat waves. This temperature increase is expected to cause a decline in kelp forest ecosystems across California. Kelp forests are needed economically by industries such as fisheries and a loss of these ecosystems would devastate California's economy and communities. To determine the future of California's kelp forests, we aimed to create a model based on existing data in this ecosystem. I acquired data on bottom sea temperatures from the Santa Barbara Coastal Long-term Ecological Research program and fish species abundance from the National Park Service (NPS). Once I accumulated and organized the data, I used Amazon Forecast's AutoPredictor to predict the future of California's kelp forests between the years of January 1st, 2017, to January 1st, 2020 for the temperature data and September 1st, 2003 to June 1st, 2005 for the fish species data. I then evaluated the accuracy of the model and cross-validated it with existing data for the time periods. From this, I found that the model's 10-percentile data, 50-percentile data, and 90-percentile data had accurate forecasts for the lower, middle, and upper parts of the data respectively. However, the model couldn't predict an accurate and consistent amount of a species of fish due to either the non-daily data or due to the acquisition of data that had many extraneous variables that the algorithm couldn't predict. From this it can be concluded that the algorithm works best for forecasting the parts of a kelp forest that have large amounts of consistent data and that the model struggles with observational data with extraneous factors.

Introduction

Kelp forests are known to have some of the most lively and thriving ecosystems in the world—a vast array of wildlife life with each being connected to another. These consist of large brown algae typically around the rocky coasts of the Western coast of the United States from Alaska and Canada to Baja California. Kelp is an organism with a rapid growth rate resulting in quick recoveries from disastrous events. Much of this is from the fact they contain pneumatocysts, which contain air bladders that help them float to the surface and get more sunlight than other plants typically can (NOAA Fisheries).

Kelp forests' importance to the ecosystem comes from their growing nature and how they have adapted to their environment. Since most ecosystems reside in rocky areas, they tend to submerge themselves under the rocks to protect themselves from the sea currents threatening their survival. Because of this, the area which was previously inhabitable becomes an area of biodiversity as the kelp slows the currents down and provides a refuge for a gamut of fish and species to live (NOAA Fisheries). Additionally, these ecosystems are essential to keeping the soil intact and decreasing the amount of soil erosion (Tait 2019).

Even though many do not think about it, kelp is an important product used by many people for reasons such as extraction of alginic acid, a source of scientific research, and a source of heritage. They are key to the California abalone industry valued at $44 million and are key to other species of fish sold commercially. To add on, they help generate markets for local people so that they can sustain themselves through tourism whether it is selling goods, giving tours, or other activities (University of California, Davis). Kelp forests are important to many lives and are an important industry for California.

Despite their ecological and economic importance, there are many current threats to kelp forests. Even though kelp's growth rate is fast, forests have seen a decrease in size, which is caused by many factors, most notably global warming (Mcpherson). Global warming is important as when the sea's temperature rises, kelp undergoes a faster decomposition process since its tissue becomes weaker. This leads to positive loop cycles as the kelp forests are not able to store carbon long enough to sink to a depth deep enough. This causes global warming to only speed up as kelp cannot convert too much carbon into oxygen resulting in a faster increase in climate (Filbee-Dexter et al. 2019). Another threat that kelp forests face is overfishing. Overfishing has resulted in notable vertebrate apex predators' numbers decreasing or disappearing and allowing for herbivorous organisms such as urchins to grow out of proportion. This results in miles of urchin-caused deforestation resulting in entire ecosystems disappearing as the kelp starts to perish (Steneck et al. 2002). Fishing regulations have helped in the deforestation of kelp forests, but losses of life such as Steller's Sea cow are still happening.

Currently, the future of kelp forests is difficult to predict due to the number of variables determining just how healthy a kelp ecosystem is. Recorded research only began in 1945 when H.L. Andrews conducted his expeditions in South America (Schiel and Foster et al. 2015). However, long-term monitoring data by agencies including information such as the sea bottom temperature, number of fish, and the substrate have provided large amounts of data that could be used by machine learning (ML) algorithms to help make predictions (Zhou et al. 2017). Additionally, ML has become incorporated in other fields, such as in medicine to look for possible early signs of diseases. This is similar to how it could be used for predicting the future of kelp forest ecosystems.

Forecasting is a type of machine learning similar to typical prediction models, (where you input some data and the algorithm tries to match a function or a set of them in order to produce an accurate forecast) except for the fact that it operates with data that is recorded based on sequences of time. Since much data collected during the research of kelp ecosystems are based on specific time intervals, forecasting is ideal because it helps build a machine learning model based on time series data, data that is collected using intervals of time.

Amazon Forecasting is a forecasting service that is an easy and accessible tool for those who are not machine learning professionals. Additionally, Amazon Forecast contains machine learning operations that compare the results of multiple different algorithms in order to provide the dataset with the ideal match for accuracy (Amazon). This could be useful as it could still operate in cases where data points are missing which could happen if some issue arises while a research expedition is operating and data is not able to be collected for logistical or safety reasons. This is a common issue with field-collected data.

AutoPredictor is a wrapper, a program that can test multiple combinations of algorithms to determine the optimal one, that is used for time series data sets and helps provide an ideal match to create accurate predictions. AutoPredictor goes through all of the different algorithms (mainly Autoregressive Integrated Moving Average (ARIMA), Prophet, and CNN-QR) that are used for time series data and tests the model. It looks based on accuracy and using this knowledge it chooses the combination of algorithms to create a final model for your personalized dataset. The models have varying ranges of statistical and mathematical operations being performed and when combined some are either more suitable or less.

Materials & Methods

For the temperature data, I used data from the Santa Barbara Coastal Long Term Ecological Research (SBC LTER) on reef bottom temperatures. The SBC LTER was established in 2000 to examine temporal and spatial patterns in giant kelp forests. This dataset includes water temperature at nine reef sites along the coast of the Santa Barbara Channel and on the north side of Santa Cruz Island. The data is from two sample temperature loggers that sample every 30 minutes, with an offset of 15 minutes to allow for data to be recorded every 15 minutes at each site. I took data from 2002 to 2016 and put it into Microsoft Excel (google sheets would also work fine. I created a new file for each new site at which the data was collected, but since some data was missing from some sites, I combined them to create a complete and continuous stream of data. However, I verified that they were close to each other so it would still be an accurate measure of data. Once I restructured the data to match the format (site, then date, then temperature each having its own column), I was able to put it into Amazon Forecast (which compared the algorithms Autoregressive Integrated Moving Average (ARIMA), Prophet, and CNN-QR and picked the best one) and picked the 50-percentile forecast and compared the results to the actual data from 2017-2019.

For the fish data, it was relatively the same process as for getting and putting the data (I used the data from the years 1985 to 2002) but there was much more data as there were 31 species along with the sites. For each, I had to reorganize the data to match the temperature data format. The process of reorganization resulted in each species having to be recategorized in the date in the first column, the station in the second, its latitude and longitude in the third and fourth respectively, the species in the fifth, and the amount in the sixth. After inputting it into Amazon Forecast, I chose the percentile forecasts and compared them with the data from 2003-2005 in quarterly forecasts. Next, I noticed that for some species there was not enough data and the percentile estimates all showed 0, and thus I only took species with at least 5 dates where the predictions have enough data. Then I removed the rows that did not have enough data and conducted the analysis.

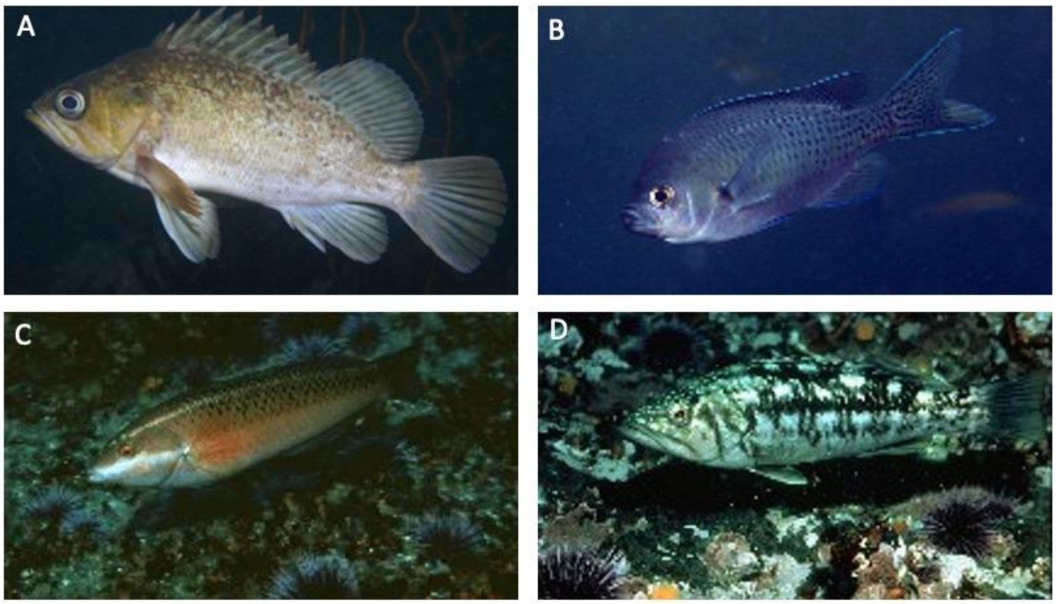

Figure 1. Representative fish species from the underwater visual transect surveys conducted by the National Park Service. (A) *Sebastes atrovirens* (kelp rockfish), (B) *Chromis punctipinnis* (blacksmith chromis), (C) *Halichoeres semicinctus* (rock wrasse), and (D) *Paralabrax clathratus* (kelp bass). Images provided by FishBase and and the SIMoN species database.

Results

Figure 2. Comparison of 10-percentile prediction of the forecast of kelp ecosystem bottom water temperature with actual kelp ecosystem bottom water temperature near the Santa Barbara

Channel Islands

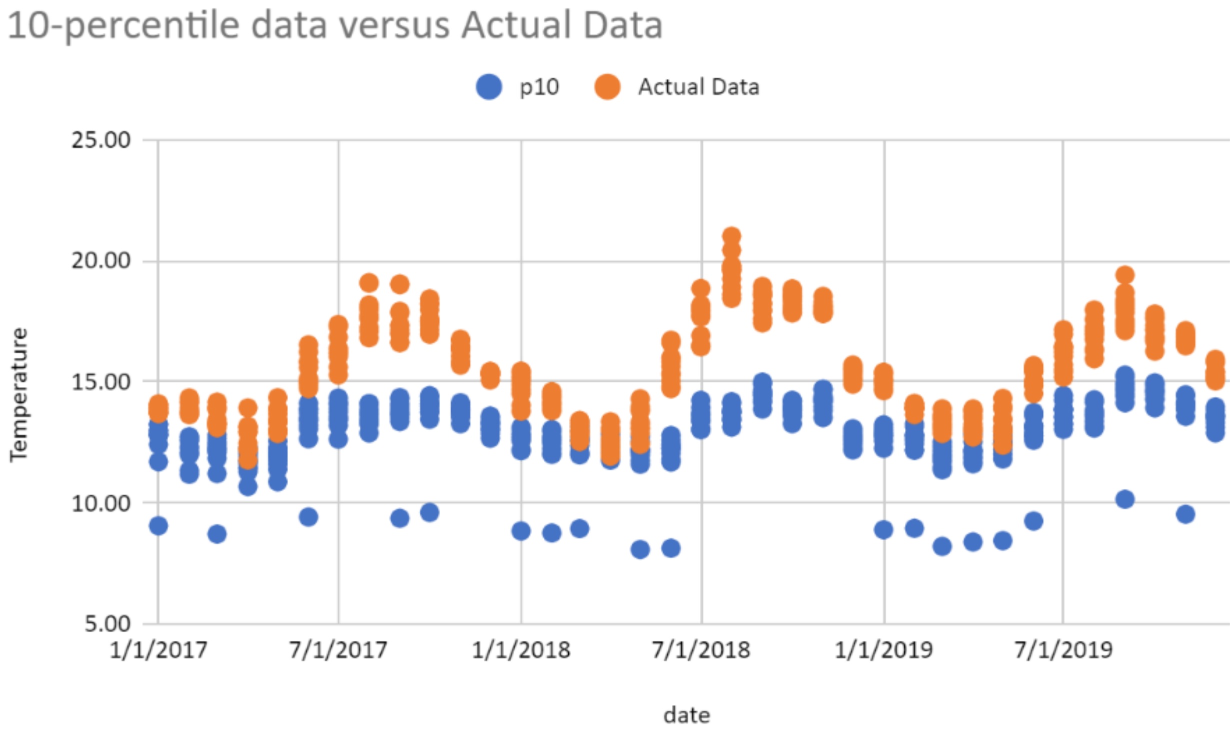

Figure 3. Comparison of 90-percentile prediction of the forecast of kelp ecosystem bottom water temperature with actual kelp ecosystem bottom water temperature near the Santa Barbara

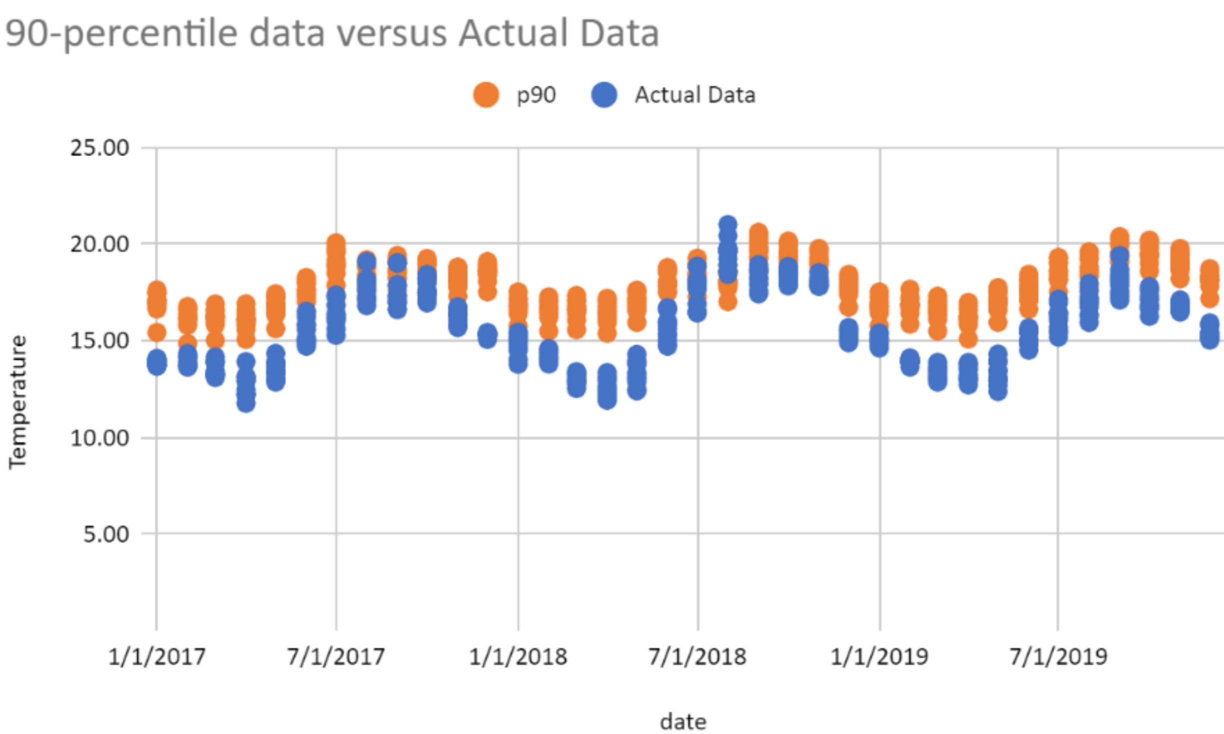

Figure 4. Comparison of 50-percentile prediction of the forecast of kelp ecosystem bottom water temperature with actual kelp ecosystem bottom water temperature near the Santa Barbara

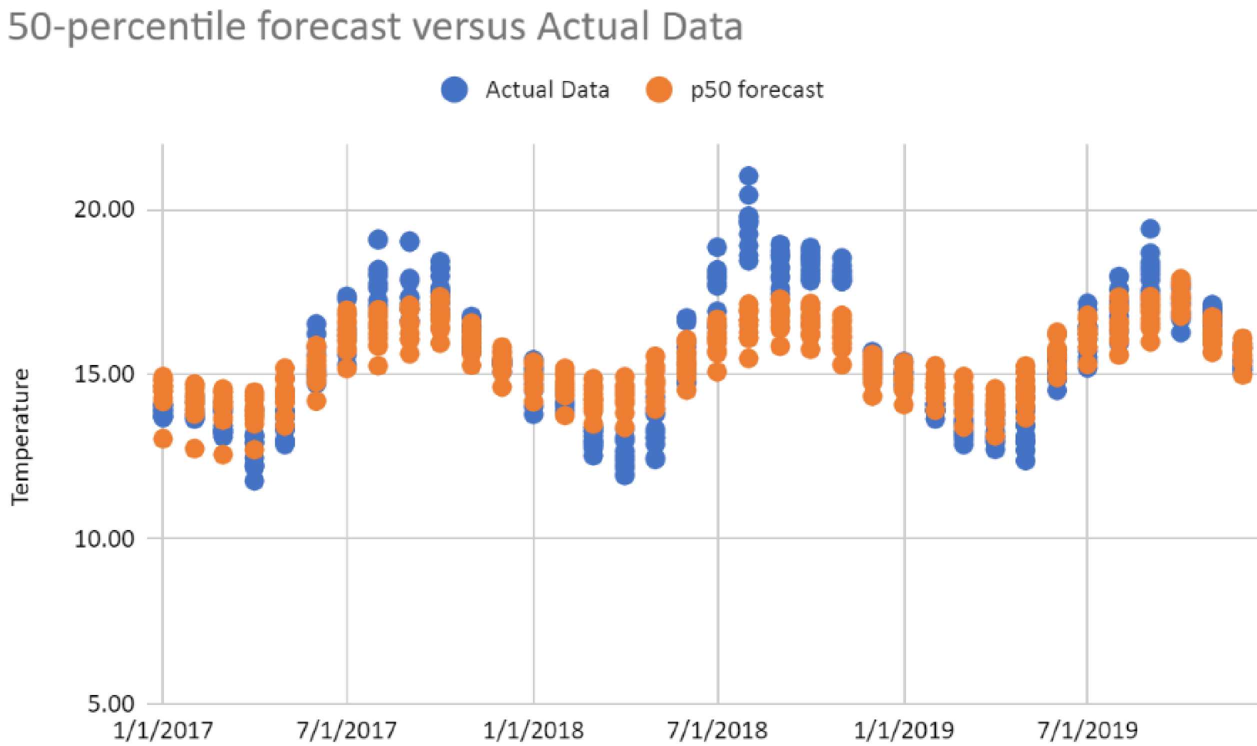

Table 1. *Chromis punctipinnis* (juvenile)

Date	p10	p50	p90	mean	Actual data
9/1/2003	-23.04	96.09	840.72	223.20	0
12/1/2003	-1.40	-1.15	-0.88	15.09	0
9/1/2004	-4.40	84.81	644.04	179.97	95
12/1/2004	-4.56	-4.25	-3.32	-2.65	0
6/1/2005	-8.13	39.15	508.27	133.00	0

Table 2. *Sebastes atrovirens* (juvenile)

Date	p10	p50	p90	mean	Actual data
9/1/2003	-0.01	0.00	5.03	1.19	12
12/1/2003	-0.02	-0.02	-0.01	0.06	9
9/1/2004	-0.04	-0.04	5.37	0.96	1
12/1/2004	0.02	0.02	0.02	0.03	0
6/1/2005	-0.18	0.07	2.32	0.47	0

Table 3. *Halichoeres semicinctus* (male)

Date	p10	p50	p90	mean	Actual data
9/1/2003	-1.29	1.20	4.67	1.46	4
12/1/2003	-2.10	0.30	2.64	0.32	1
9/1/2004	-1.62	1.06	5.29	1.52	1
12/1/2004	-2.22	0.39	2.89	0.37	0
6/1/2005	-0.60	3.09	7.66	3.12	0

Table 4. *Paralabrax clathratus* (adult)

Date	p10	p50	p90	mean	Actual data
9/1/2003	-3.67	8.57	21.12	8.76	4
12/1/2003	-11.70	-2.94	5.61	-2.75	1
9/1/2004	-8.97	3.74	17.41	4.44	1
12/1/2004	-12.46	-3.23	5.63	-3.20	0
6/1/2005	-3.36	13.04	31.89	12.91	0

Table 5. *Chromis punctipinnis* (adult)

Date	p10	p50	p90	mean	Actual data
9/1/2003	0.00	211.00	633.00	272.75	843
6/1/2004	0.00	681.00	1595.00	709.56	544
9/1/2004	0.00	254.00	633.00	310.47	60
3/1/2005	0.00	0.00	10.00	53.82	250
6/1/2005	0.00	649.00	1125.00	673.91	955

Table 6. *Halichoeres semicinctus* (female)

Date	p10	p50	p90	mean	Actual data
9/1/2003	-2.24	2.30	9.27	2.88	2
12/1/2003	-3.85	0.62	4.99	0.67	1
9/1/2004	-2.33	2.59	10.18	3.53	7
12/1/2004	-4.02	0.74	5.30	0.70	0
6/1/2005	-0.47	6.51	15.15	6.68	0

Figure 2 shows a comparison of the 50-percentile forecast of the bottom temperature data and the actual data and here we see a unique trend. The data which is forecasted typically covers the middle section of the data. Additionally, it on average has a deviation of 0.987 degrees making it seem as if the graph is great for forecasting the temperatures when the ocean is not at its peak such as from the months from August to September and January to February (as shown by Fig. 2). However, there are issues when the forecast is compared to the extremes of the observational dataset. Also, the average difference between the real data and the forecast is 0.02 degrees which shows that the forecast is close to the average. The 50-percentile forecast does not accurately capture the extremes of the temperature variations, preventing accurate estimates about the higher data and lower data which would occur when oceans are typically warmer or colder during the summer and winter months respectively.

Comparing the 50-percentile forecast (Figure 3) to the 10-percentile forecast, we see that the algorithm is able to much more accurately predict the temperature of the sea when it is colder than average during the months of January to February. This is useful as this would be much more efficient for collecting seasonal data. However, unlike the 50-percentile forecast, it has made bad estimates when the temperature is around the average with a deviation of about 2.654°C. This is a considerable deviation since the temperature is measured in °C. Additionally, it has an average of forecasting -2.52°C, which is an indicator that it typically forecasts data that is less than the true data.

Neither of the above forecasts account for the warmer months from August to September, but the 90-percentile forecast (Figure 4) does so accurately. It has estimates which are typically higher than the true data as seen with their typical deviation from the real-time data being 2.418°C. At first, it may seem problematic, but as shown in the last graph, these predictions much more accurately plot to forecast predictions with a higher prediction and this is proved by the fact that on average it predicts 2.33°C higher than the real data.

From the predictions generated with the fish abundance data, we see it is tough to find a comparison. The predictions were made on three monthly scales but for some (especially from December to March (1/12 were present) and March to June (5/12 were present)) there was not enough data. Additionally, there are issues with the data as 24/30 of the 10 percentile predictions resulted in a negative number which is not possible. To add on, the ones where we could see some close predictions to the actual data were for fishes such as female *Halichoeres semicinctus* (rock wrasse) in Table 6, male *Halichoeres semicinctus* (rock wrasse) in Table 3, and juvenile *Sebastes atrovirens* (kelp rockfish) in Table 2. There were no consistent predictions for any of the percentiles (10-, 50-, or 90th percentiles) for all the species, such that there was no ideal percentile. Additionally, these predictions were more accurate for these types of fish because the abundances from the underwater visual surveys were typically small, especially compared to other species like the juvenile *Chromis punctipinnis* (Blacksmith chromis) where the number of individuals much greater. This suggests that the model provides more accurate predictions of species abundances for fishes with low starting abundance from the underwater surveys. Additionally, the model doesn't consider the variability inherent in estimating fish abundances which goes into generating population estimates for fish. To add on, the lack of consistent data makes it tough for the model to come up with accurate predictions as it only has periodic samples with large gaps between them. All this contributes to the inaccuracy of the model when it predicts fish.

Discussion:

From the above analysis, I come to the conclusion that using machine learning can be an effective method to predict the future of kelp ecosystems, with some important caveats.

In the case of predicting factors for which we can acquire a lot of data and are generally more predictable (e.g., abiotic variables), machine learning is an effective tool that can be used for predicting the future. However, it is still recommended to include a “human-in-the-loop” approach especially when calculating temperature because the program treated all data equally and can run into problems when it must interpret data based on specific conditions which computers do not know. Thus, the best plan for answering this would be to combine parts of each type of forecast (the lower third as the 10-percentile model’s prediction, the middle third as the 50-percentile model’s prediction, and the last third as the 90-percentile model’s prediction). This would allow for the lower values to be accurate with the extremes due to two models being accurate in those measures and the average of the two extreme forecasts being able to accurately predict the middle part of the model. However, I note that there was a lot of data needed to reach this result and even then, there still may be some overtly wrong estimates, so ideally if there is a seasonal or a pattern that can be predicted, picking the forecast which most associates with that pattern would be ideal (e.g., the 10 percentile model being used when the weather will be cold and 90-percentile model when it is predicted to be hot).

However, the opposite could be said regarding the ability of the algorithm to predict fish abundance. There was a small chance that the model would be able to accurately forecast the given number of fish and there were many factors contributing to this. One was that of non-consistent data. Unlike the temperature data collected, the fish data were collected once or twice every year at a location resulting in data not being present in parts of the year such as between December and March resulting in not accurate predictions in those areas. To add on, the model does not understand the chance of sometimes transects having large amounts or small amounts and this leads it to see it as a plausible data point and thus not being able to predict the species amount accurately. Not to include that it took decimals and negatives for a population count. These are only some factors that contribute to the model's ineffective forecasting, and from this, I conclude that machine learning cannot be used for predicting everything. This is true when data is not collected in a consistent format because it causes the algorithm to not be able to make predictions for that time of the year. Additionally, it doesn't understand human-retained knowledge such as population cannot be negative or a decimal because it sees all these numbers as simply a number rather than a population number which has to be a non-negative integer. Furthermore, it has a poor time considering anomalies in the data or chance and probability which causes its prediction to become far less accurate. Some solutions are a larger amount of data as it may have become better at diluting the impact of anomalies, and further human supervision and input to the models.

Conclusion

Machine Learning would be a good solution for predicting the future of kelp forests, especially when it comes to forecasting variables that have a large amount of consistently acquired data with less uncertainty, but otherwise, it may not result in accurate predictions. For this project in particular, I found that the model was better at predicting abiotic variables (e.g., temperature) with high accuracy and precision, but requires improvements to be useful for the prediction of biotic variables (e.g., fish species abundance from underwater transect surveys).

Acknowledgements

I would like to thank the Santa Barbara Coastal LTER and the National Park Service for collecting and providing the data on kelp forest bottom temperatures and fish species abundance. I would also like to thank Dr. Beverly French from the Scripps Institution of Oceanography at UC San Diego for guiding me through the research project.

References

[1] Blechman, N. (2008). *Forecast*. Amazon Forecast Algorithms. Retrieved April 29, 2023, from https://docs.aws.amazon.com/forecast/latest/dg/aws-forecast-choosing-recipes.html

[2] *Chromis punctipinnis summary page*. FishBase. (n.d.-a). https://fishbase.mnhn.fr/summary/3646

[3] Filbee-Dexter, K., Wernberg, T., Fredriksen, S., Norderhaug, K. M., & Pedersen, M. F. (2019). Arctic kelp forests: Diversity, resilience and future. *Global and Planetary Change*, *172*, 1–14. https://doi.org/10.1016/j.gloplacha.2018.09.005

[4] Fisheries, N. O. A. A. (2021, November 2). *Kelp Forest Habitat on the West Coast*. NOAA. Retrieved April 23, 2023, from https://www.fisheries.noaa.gov/west-coast/habitat-conservation/kelp-forest-habitat-west-coast

[5] *Halichoeres semicinctus Summary Page*. FishBase. (n.d.-b). https://www.fishbase.se/summary/Halichoeres-semicinctus.html

[6] McPherson, M. L., Finger, D. J. I., Houskeeper, H. F., Bell, T. W., Carr, M. H., Rogers-Bennett, L., & Kudela, R. M. (2021, March 5). *Large-scale shift in the structure of a kelp forest ecosystem co-occurs with an epizootic and marine heatwave*. Nature News. Retrieved April 23, 2023, from https://www.nature.com/articles/s42003-021-01827-6

[7] *Paralabrax Clathratus summary page*. FishBase. (n.d.-c). https://www.fishbase.se/summary/3335

[8] Santa Barbara Coastal LTER, D. Reed, and R. Miller. 2022. SBC LTER: Reef: Bottom Temperature:
Continuous water temperature, ongoing since 2000 ver 28. Environmental Data Initiative. https://doi.org/10.6073/pasta/734d6826266051b44be73f9e816b5103 (Accessed 2023-04-23)

[9] Schiel, D. R., & Foster, M. S. (2015). Introduction to giant kelp forests worldwide. *The Biology and Ecology of Giant Kelp Forests*, 3–22. https://doi.org/10.1525/california/9780520278868.003.0001

[10] *Species database*. SIMoN. (n.d.). https://sanctuarysimon.org/dbtools/species-database/id/208/sebastes/atrovirens/kelp-rockfish/

[11] Steneck, R. S., Graham, M. H., Bourque, B. J., Corbett, D., Erlandson, J. M., Estes, J. A., & Tegner, M. J. (2002). Kelp forest ecosystems: Biodiversity, stability, resilience and future. *Environmental Conservation*, *29*(4), 436–459. https://doi.org/10.1017/s0376892902000322

[12] Tait, L. W. (2019). Giant kelp forests at critical light thresholds show compromised ecological resilience to environmental and biological drivers. *Estuarine, Coastal and Shelf Science*, *219*, 231–241. https://doi.org/10.1016/j.ecss.2019.02.026

[13] Troell, M., Robertson-Andersson, D., Anderson, R. J., Bolton, J. J., Maneveldt, G., Halling, C., & Probyn, T. (2006). Abalone farming in South Africa: An overview with perspectives on kelp resources, abalone feed, potential for on-farm seaweed production and socio-economic importance. *Aquaculture*, *257*(1-4), 266–281. https://doi.org/10.1016/j.aquaculture.2006.02.066

[14] Vásquez, J. A., Zuñiga, S., Tala, F., Piaget, N., Rodríguez, D. C., & Vega, J. M. (2013). Economic valuation of kelp forests in northern Chile: Values of goods and services of the ecosystem. *Journal of Applied Phycology*, *26*(2), 1081–1088. https://doi.org/10.1007/s10811-013-0173-6

[15] Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on Big Data: Opportunities and challenges. *Neurocomputing*, *237*, 350–361. https://doi.org/10.1016/j.neucom.2017.01.026