2024 Impacts of Sea Level Rise Challenge

Penguins

Victoria Cicherski, Anjana Mittal

Executive Summary

Our report details the use of advanced machine-learning methods and predictive analytics to prevent and reduce the consequences of sea level rise on a community and environment. An efficient approach to minimizing the harmful effects of sea level rise is through prediction. To achieve this, we utilized sea level change and different types of flooding data for California from multiple sources. Then, we analyzed our dataset and implemented many machine-learning algorithms until we felt satisfied with the model's performance. The insights gained from our predictive modeling and analysis can aid in preventative strategies such as the protection of critical infrastructure and the allocation of resources, as well as assessing the damage to the Bay Area community.

Problem Statement

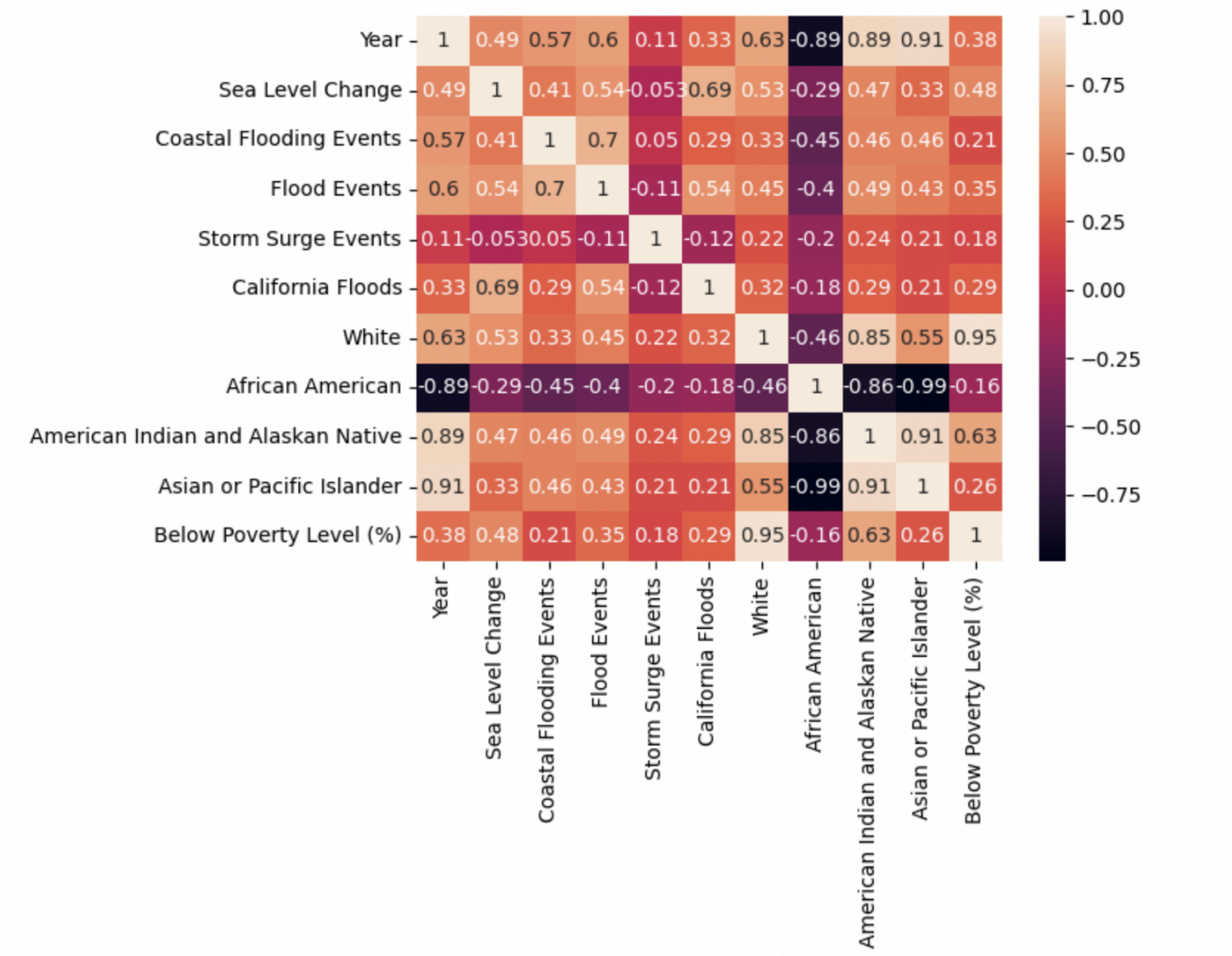
Sea level rise is influenced by an intricate interplay of factors that are unique to each part of the world. Our project seeks to address the impact of sea level rise within San Francisco and how it influences the community and environment of the Bay Area.

Datasets

For our analysis of the surrounding area of San Francisco, we used multiple datasets from various sources: the "Annual mean sea level trends" dataset from OEHHA (California Office of Environmental Health Hazard Assessment), the "Historical Flood Impact" dataset from FEMA (Federal Emergency Management Agency), as well as the 1990, 2000, and 2010 census datasets for Marin and San Francisco County from Bay Area Census. We compiled all of the data into a single CSV file which we called "Sea-Level-Change-Impacts."

Data Exploration

The first step of our data analysis was to create graphs to examine how sea level rise within San Francisco influenced natural events such as flooding and the demographics of the Bay Area community. Our next step was to create a correlation matrix.



From the matrix, we could see that all of the variables except for storm surge events and the African-American population have a good correlation with sea level change. The correlation matrix enabled us to understand which features would help predict sea level rise impact. Although the demographic data within the Bay Area had mostly positive correlations with sea level change, we decided to not employ the data for our machine learning models based on logical reasoning. Since we pulled the data from the US Census which occurs every ten years, we were only able to use three censuses. Thus, we determined that the demographic data would help in understanding the impacted community but would not be favorable in predicting sea level rise impact.

Methodology

Because our goal was to predict a discrete quantity, we used regression analysis to identify the relationship between sea level rise and flooding events. We used many different models to determine which best illustrates the relationship. We used multivariate regression, simple linear regression, decision trees, random forests, logistic regression, and polynomial regression models to find the model of best fit.

Modeling and Analysis

Since we were trying to predict a discrete quantity, our approach was to explore different regression modeling methods. We split our dataset into training, validation, and testing data for each model we created. We began with a multivariate regression model where our independent variable was sea level change and our dependent variables were coastal flooding, flooding events, and overall flooding events in California. The result of the testing data was a low R-squared value and an MSE of 265.

Next, we created a simple linear regression model with sea level change as the independent variable and coastal flooding events as the dependent variable. After fitting the model, the validation data produced an MSE of 0.44 and an R-squared value of 0.49. However, when we evaluated the model on test data, the MSE was 8218 and the R-squared value was -5.9.

We then move on to multiple linear regression and fit a model with sea level change and California floods as the predictor variables and coastal flooding events as the dependent variable. Here, the validation data produced an MSE of 0.45 and an R-squared of 0.49.

Next, we move on to decision trees. For our first attempt, sea level change is the independent variable and coastal flooding events, flood events, storm surge events and California floods are the dependent variables. The MSE for the validation data is 475.2 and the R-squared is -0.6. Overall, not a good model fit. We then decided to switch the independent and dependent variables, meaning sea level change is the dependent and all the events are the predictor variables. We found that the MSE for validation is 469 and the R-squared is 0.48, a surprising result. We again use decision trees with sea level change as an independent variable and just flood events as the dependent variable. This is not a good model, with a validation MSE of 66 and the validation R-squared of -1.

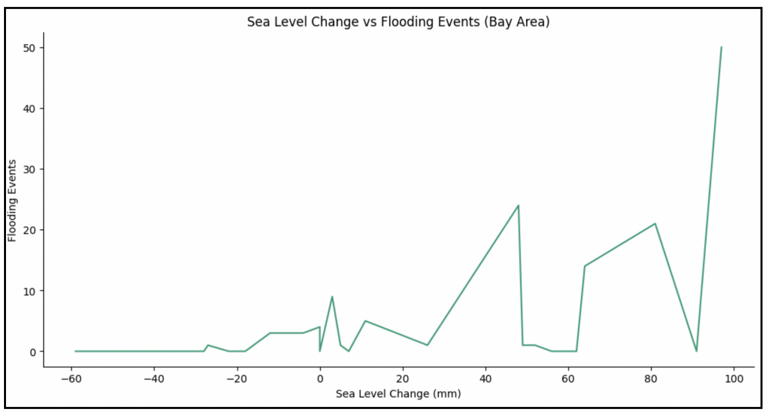
We move on to random forest modeling. For our first attempt, sea level change is the independent variable and coastal flooding events, flood events, storm surge events and California floods are the dependent variables. The MSE is 240 and the R-squared is -0.23, not a good model. Switching the dependent and independent variables around, as done with decision trees, we find a MSE of 214 and a R-squared value of 0.76, again pointing toward the same relationship as the decision tree model. Finally, using sea level change, flood events and coastal flooding events as the predictor variables and California floods as the dependent variable, we fit yet another random forest model. The MSE is 1782 and the R-squared is -0.49, showing another poor model fit.

Next, we try a logistic regression model with sea level change as the independent variable and coastal flooding events, flood events, storm surge events and California floods as the predictors. This model produces a 99 MSE value and a 0.89 R-squared value, pointing towards a strong relationship.

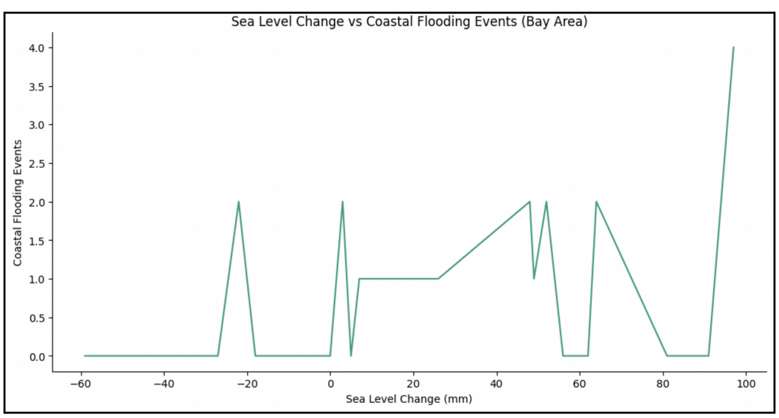
Lastly, we use polynomial regression to express the relationship between these variables. Here, we decided to stop splitting into train, test and validation data sets and just split into train and test sets. For our first model, sea level change, coastal flooding events and flood events are the predictor variables in this case and the dependent variable is California floods. This produces a high MSE of 18930 and a R-squared of -3.9. For our final model, we attempt to use sea level change and coastal flooding events to predict flooding events. Here, the MSE is 4.21 and the R-squared value is 0.83.

After evaluating all the models we fit, we decided the polynomial regression model with sea level change and coastal flooding events as the predictors and flood events as the dependent variable was our best model. This model has the lowest MSE value (4.21) and the highest R-squared value (0.83) while also illustrating that one of the impacts of sea level rise is flooding. However, another interesting relationship to note and one that carries across multiple model selections, is that flooding events can be used to predict sea level rise.

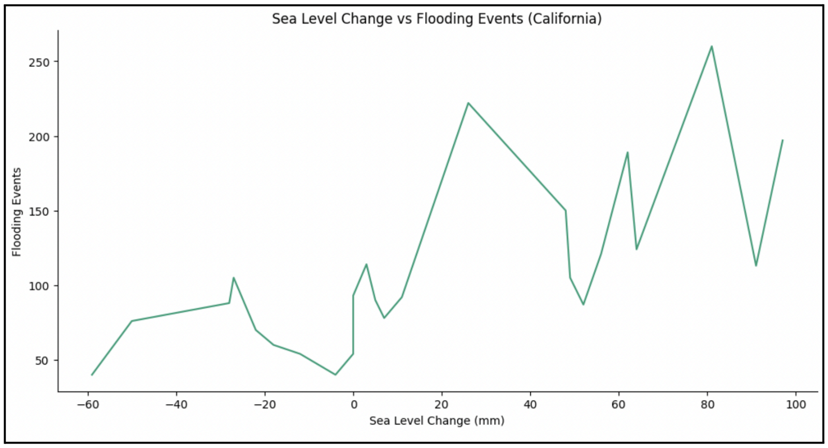
Visualization and Interpretation



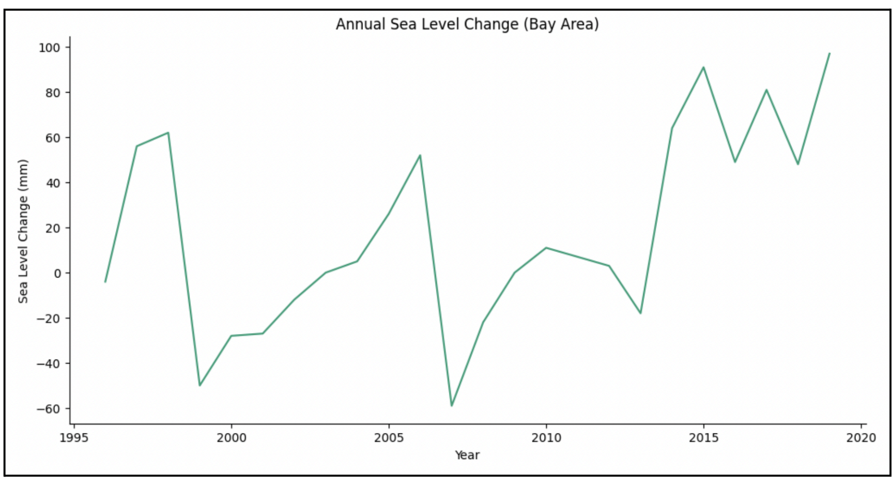
Our model revealed a somewhat increasing relationship between sea level rise and total flood events in the Bay Area. As the sea level has a positive change, there is an increase in flood events for that year. Within recent years, there has been a dramatic spike in the number of flood events.



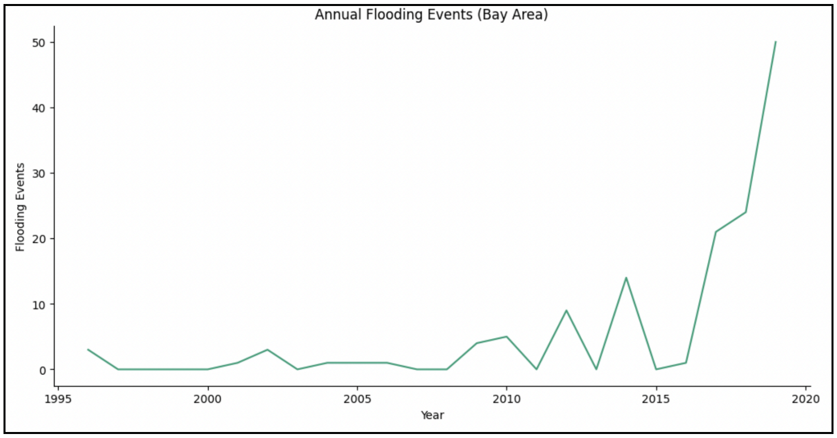
The graph, ‘Sea Level Change vs Coastal Flooding Events (Bay Area)’ shows a similar trend: as there is a positive change in sea level, the number of coastal flooding events somewhat increases. The graph also demonstrates a sudden increase in coastal flooding events in recent years.



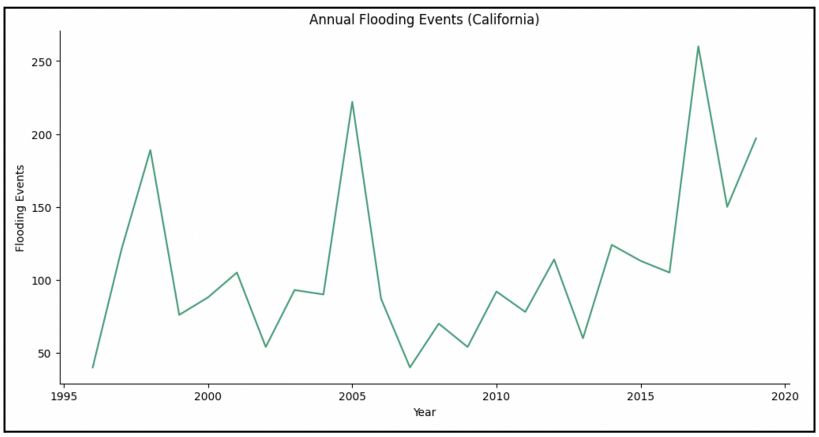
For the state of California, there is a clear positive relationship between sea level change and the number of flooding events.



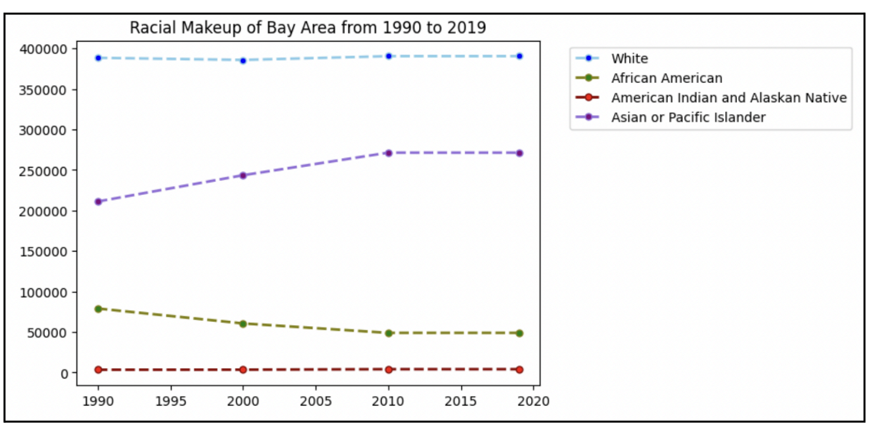
Here, the graph depicts the relative annual sea level change in millimeters for San Francisco. We can see that the sea level is increasing positively between 1996 and 2019.



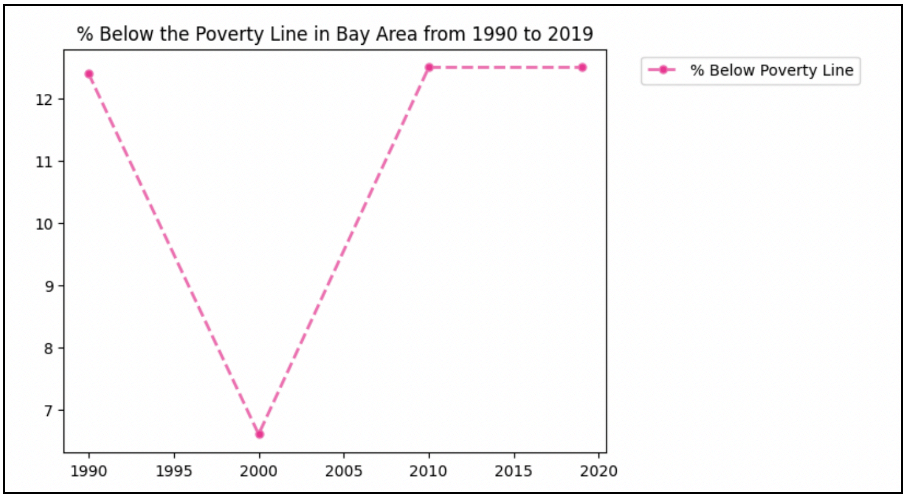
This visualization depicts the number of flooding events per year in the Bay Area. Although there is relatively stagnant growth from 1996 to 2015, we can see an obvious positive linear trend in recent years.



For the state of California, there is an unclear trend between 1996 to 2007. But afterward, there is a gradually positive increase in flooding events per year. For the next aspect of our project, we used data visualizations to understand the demographics of the Bay Area to understand who is most affected by sea level rise and the corresponding increase in flood events. We collected demographic data from the 1990, 2000, and 2010 censuses and compiled the data into line graphs for ease of understanding.



. In the visualization, ‘Racial Makeup of Bay Area from 1990 to 2019’, the populations of different racial backgrounds are illustrated throughout the years. We can see that white and Asian or Pacific Islander citizens constitute the majority of the Bay Area population.



This graph portrays the percentage of citizens living below the poverty line in the Bay Area from 1990 to 2019. The percentage of people living below the poverty line decreases after 1990, but increases after the 2000 census.

End User Communication

We created a website so that users have an interactive platform to learn more about our model and analysis.

https://janakish.wixsite.com/penguins

Conclusions and Recommendations

The rise in sea levels across the globe has unprecedented consequences on humans, the economy, and the environment; therefore, it is imperative to be able to forecast the effects of sea level rise. Our project tackles this issue through machine learning and predictive analytics to forecast the effects of sea level rise within the surrounding region of San Francisco with data compiled from the OEHHA, Bay Area Census, and FEMA. Machine learning and predictive analytics can aid in informing the placement and protection of critical infrastructure, evaluating societal and environmental impacts, and enacting policies that prioritize public health and safety.

Overall, our most effective model applied sea level change and coastal flooding events to predict the number of flood events in the Bay Area. The prediction of flood events can mitigate the harm that would ordinarily transpire with preemptive strategies. A limitation we encountered was the amount of data points that our dataset held. If we had a larger dataset to feed our models, we believe that the generalization of our models would improve. The data visualizations we created highlight the growth and change of coastal flooding and flood events that reflect the gradual rise in sea level change from 1996 to

2019. The graphs of demographic data showcase the people in the Bay Area community most impacted by the rising sea level. This information enables a more effective allocation of resources to reduce the potential damage caused by flooding in the Bay Area community. Our recommendations include building stormwater pumps, upgrading roads and bridges, constructing seawalls, and improving stormwater drainage systems. Ultimately, the outcome of our project reveals the capability of machine learning and predictive analytics in mitigating the impacts of sea level rise on the environment and society.

Sources

Oehha.ca.gov. (n.d.-b). https://oehha.ca.gov/climate-change/epic-2022/impacts-physical-systems/sea-level-rise

Historical flood risk and costs. FEMA.gov. (n.d.). https://www.fema.gov/data-visualization/historical-flood-risk-and-costs

Ncei. (n.d.). *Storm events database*. National Centers for Environmental Information. https://www.ncdc.noaa.gov/stormevents/

Census 2010 highlights - san francisco. (n.d.). https://www.sfdph.org/dph/files/reports/PolicyProcOfc/CensusBureau2010Highlights_03062012.pdf

San Francisco City and County. Bay Area Census -- San Francisco County -- 1970-1990 census data. (n.d.). http://www.bayareacensus.ca.gov/counties/SanFranciscoCounty70.htm

Marin County. Bay Area Census -- marin county. (n.d.). http://www.bayareacensus.ca.gov/counties/MarinCounty.htm