Predicting risks of toxic domoic acid events in California coast region

Toxic algal bloom

Image from www.phys.org

Toxic algal blooms are dangerous and costly

- Algal blooms in California coastal region can release domoic acid.
- Domoic acid is a neurotoxin that can cause **serious symptoms, even death**, in human.
- Domoic acid can be consumed by fish and shellfish, which makes them unsafe to consume.
- Delay/closure of fishery caused by domoic acid have costed **millions of dollars**.

Government agencies and fishing industry will be interested in prediction of such toxic events.

caseagrant.ucsd.edu

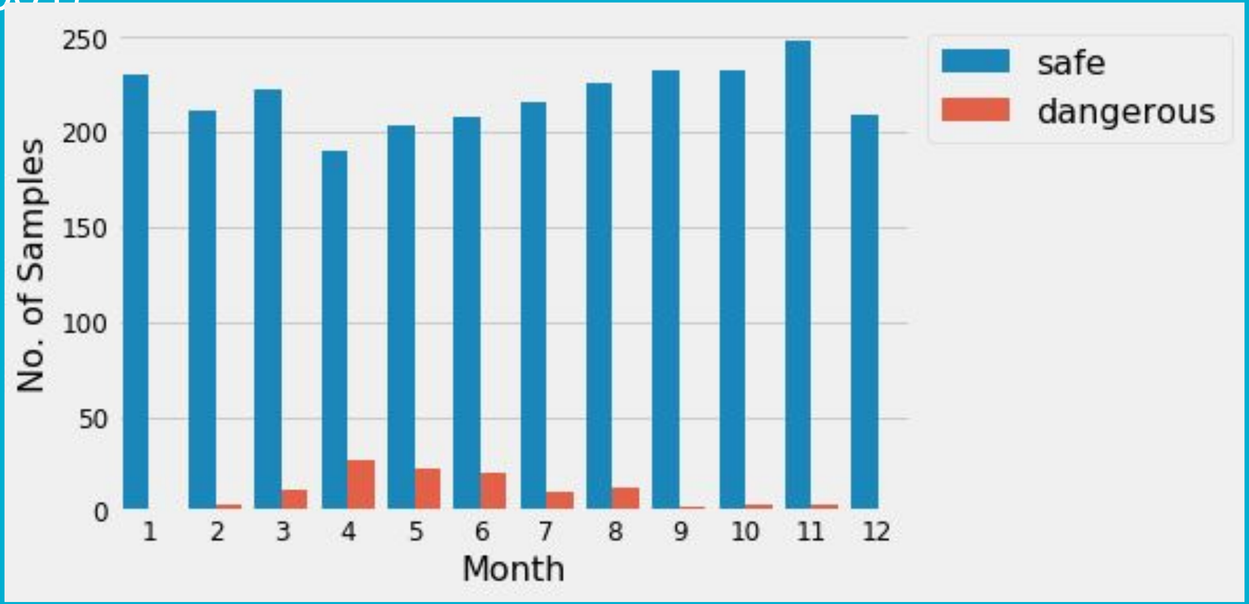
nage: Oregon Department of Fish and Wildlife

What data do we have and where are they

- Southern California Coastal Ocean Observing System provides domoic acid along with relevant physical, chemical, and biological data for ~3,500 samples.
- However, some of the data were poorly labeled. I obtained correctly labeled data from Dr. Jayme Smith, a domoic acid expert.
- Oceanography data were obtained from two other government research organizations.
- Cleaned data include 2,750 samples and 17 independent variables.

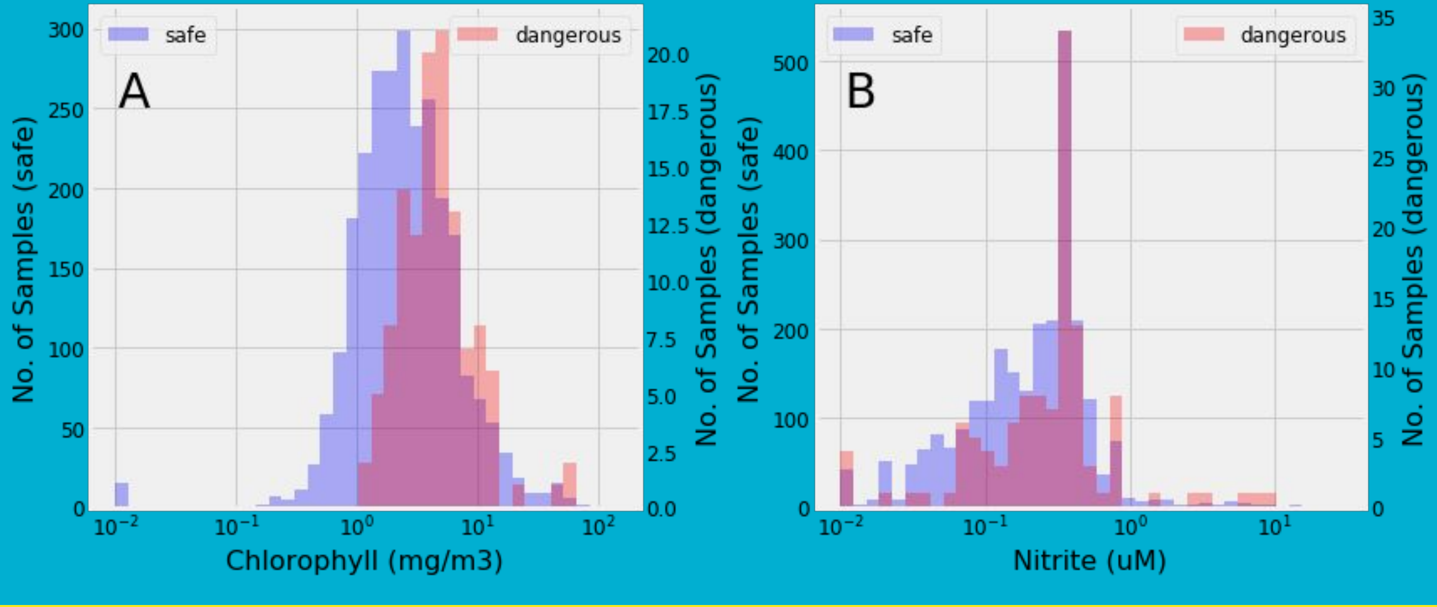
Distribution of toxic events

- Toxic events are rare. They mostly happen during spring and summer. ($p<0.001$)



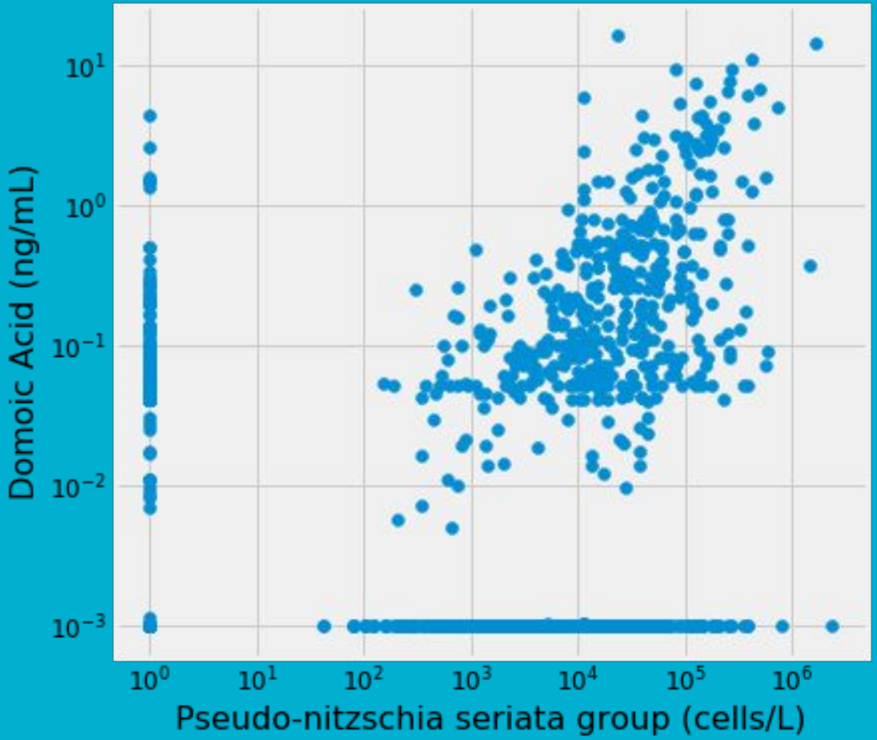
Correlation of toxic events with chemical data

- Toxic samples had higher chlorophyll ($p<0.001$) and nitrite concentrations ($p=0.006$).



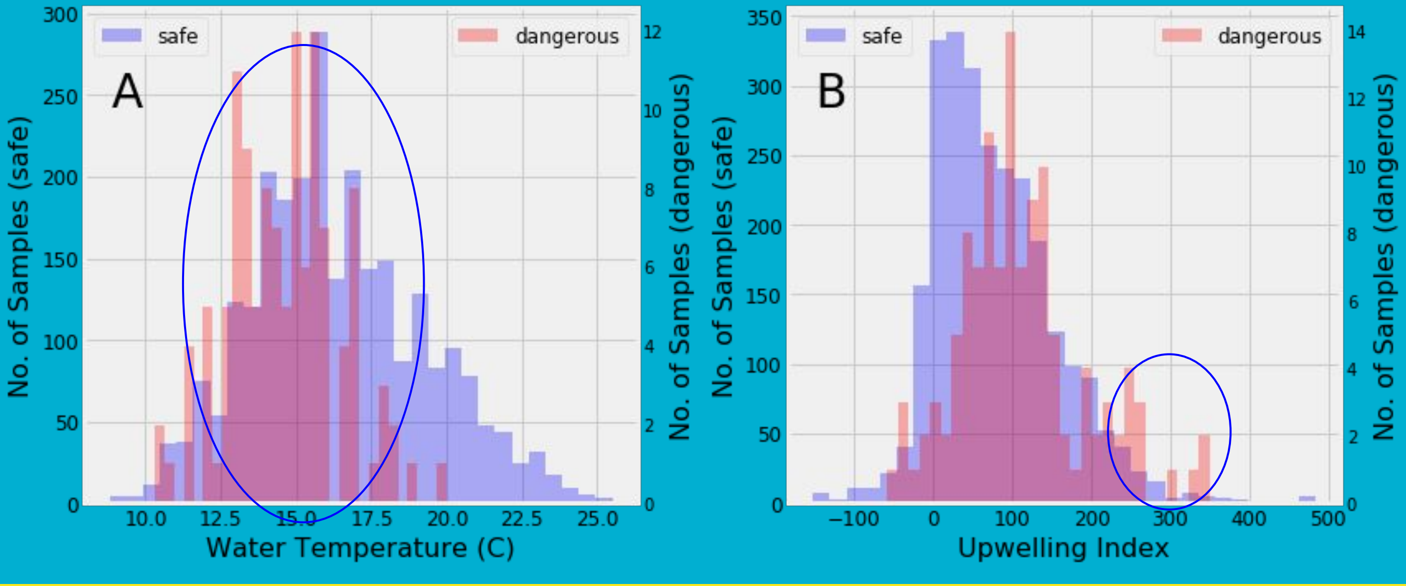
Pseudo-nitzschia produce domoic acid

- Two pseudo-nitzschia groups were found to have higher concentrations in toxic samples.
- Pseudo-nitzschia seriata and domoic acid concentrations were somewhat linearly correlated.



Correlation of toxic events with physical data

- Toxic events were more likely to occur when water temperature was between 12.5 and 17.5 degrees C, and when Ocean Upwelling Index was higher.

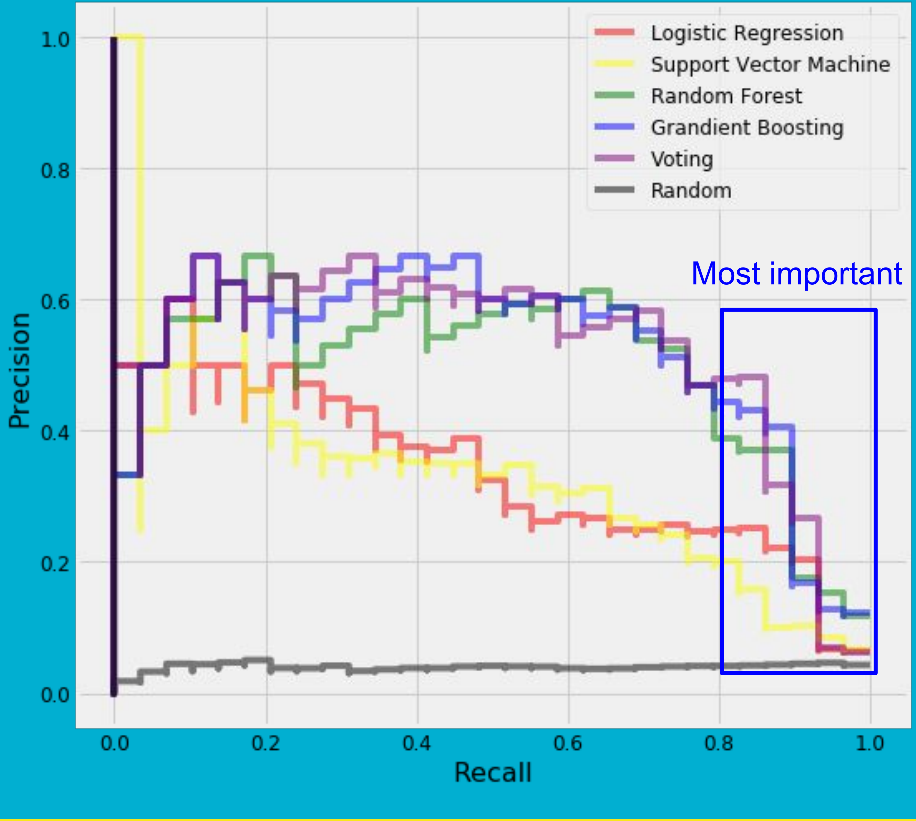


Model performance

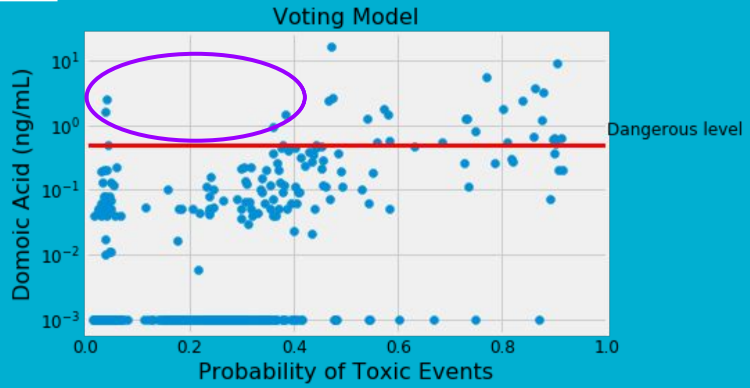
	Logistic Regression	Support Vector Machine	Random Forest	Gradient Boosting	Voting Classifier
Average precision	0.34	0.34	0.50	0.52	0.53
Accuracy	0.75	0.91	0.97	0.97	0.91
No. of true positives	27	20	20	18	22
No. of false positives	173	56	15	12	20
No. of false negatives	2	9	9	11	7
No. of true negatives	486	603	644	647	639
Precision*	0.14	0.26	0.57	0.60	0.52
Recall*	0.93	0.69	0.69	0.62	0.76

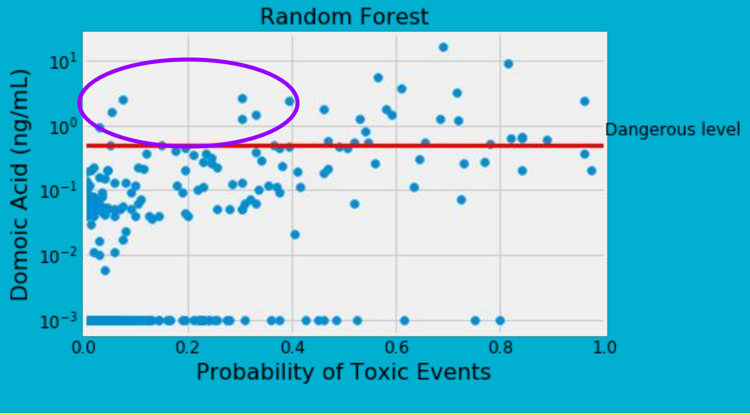
Precision and Recall

- Consequences of false positives are largely monetary (extra tests, closed beach, etc.)
- Consequences of false negatives are potentially catastrophic (public health crisis).
- Higher recall is more important.
- Voting model and Random Forest perform best at higher recall.



Predicted probability and toxin concentrations





- Toxic samples with low predicted probabilities are especially dangerous.
- Random forest model have more samples of this type.
- Overall, the voting model is the most useful.

Future directions

- Collect more data.
  - This data set was small. The number of toxic samples were especially small. More samples would help build better models.
- Predict toxic events in advance.
  - Build models to predict toxic events weeks into the future would help government agencies take proactive measures.
- Identify important features not in this study.
  - There were a few samples that none of our models got right. There must be other features that are associated with risks of toxic events. It’s important to identify them.