



# Coral Sense

Theresa Sumarta  
Tigran Poladian  
Kiara Monahan  
Ted Johnson  
Nigel Lewis

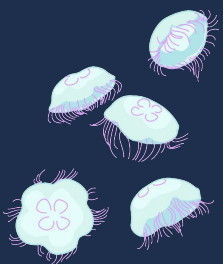


# For Most of History...

Coral reefs, often called the rainforests of the sea, once thrived across the globe, supporting rich ecosystems and providing invaluable resources to both marine life and humanity.



# A Vital Ecosystem



25%

Of all marine life is in reefs

500M

People depend on reefs

45,000 miles

Coastline protected by reefs

25%

Of fish caught in developing countries is from reefs

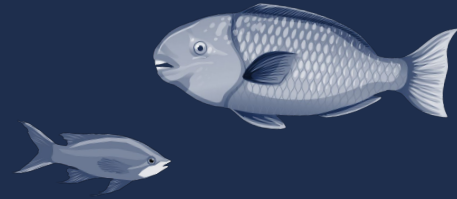
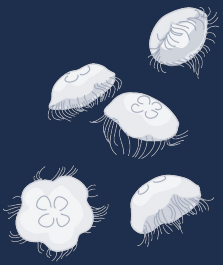


# The Decline of Coral Reefs

- Coral reefs began showing signs of significant decline as early as the 1950s
- By the mid-1990s, scientists observed widespread coral bleaching events
- The situation worsened between 2014 and 2017 during the longest global coral bleaching event in recorded history, which devastated over 75% of reefs worldwide.



# The Turning Point



Pollution



Pathogens



Overfishing



Climate  
Change

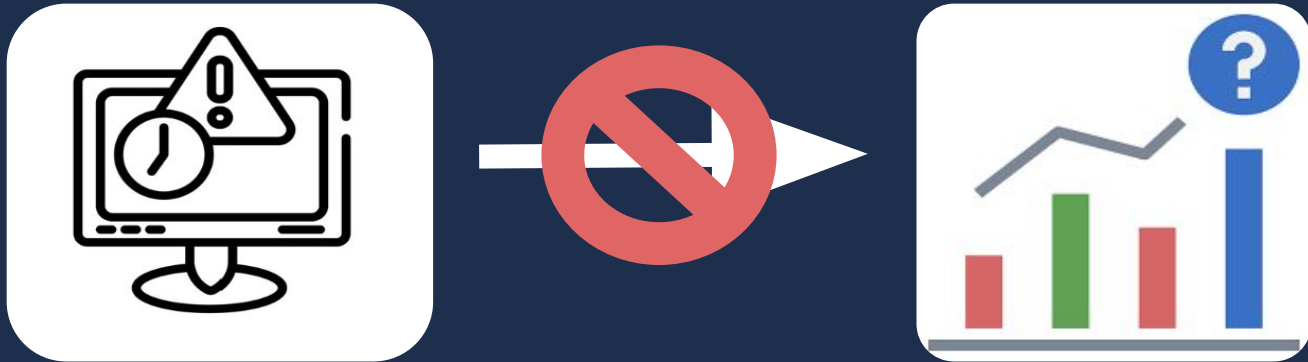


Nutrient  
Changes



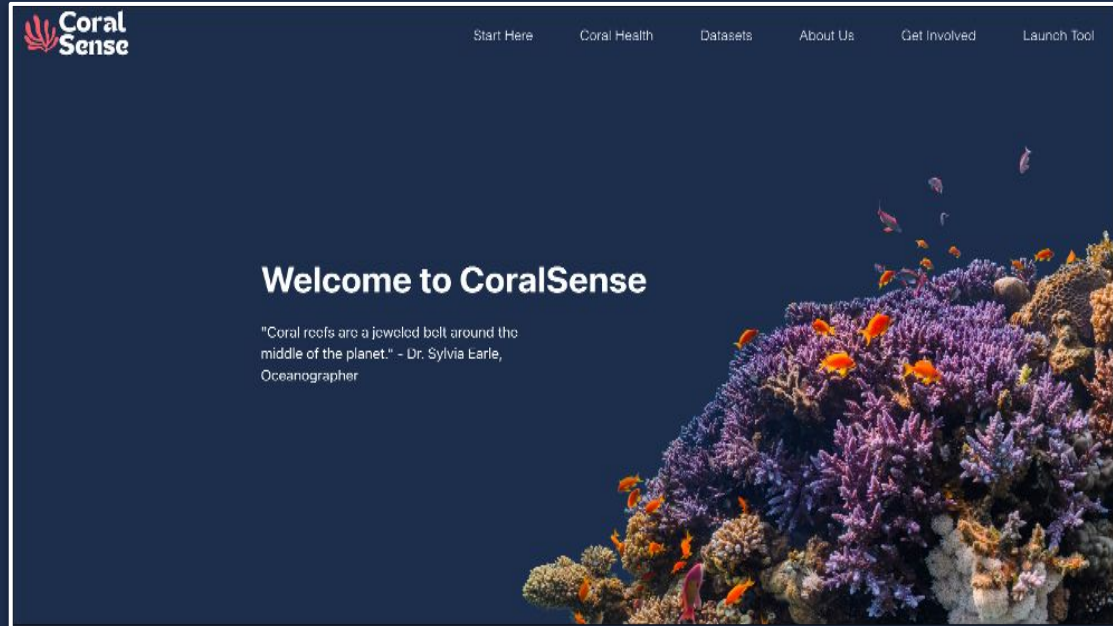
# The Call for Action

- The world calling for heroes to rise.
- Researchers, conservationists, and policymakers are seeking ways to protect what remains.
- But their tools are outdated, reactive instead of proactive, and cannot predict where to act before it is too late.



# Our Solution

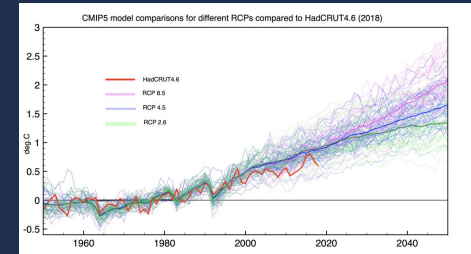
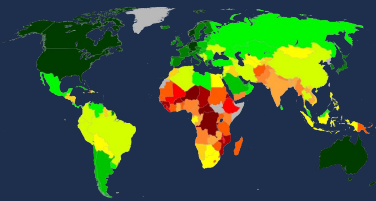
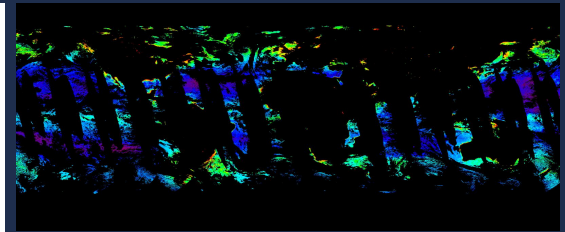
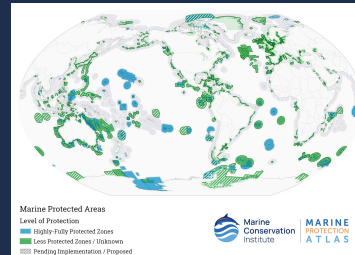
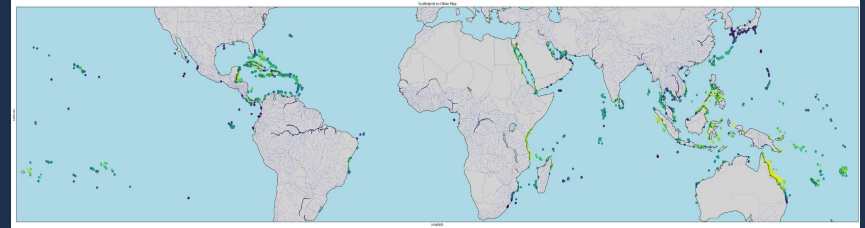
- Interactive Map interface
- Future predictions on reef health (hard coral cover) for locations around the world
- Time series predictions of coral bleaching
- Allows users to adjust key features



# Data Sources

~ 12,000 Training records

1. MODIS Aqua Level-3 Remote Sensing Reflectance (Rrs) Image Dataset
2. Global Bleaching Environmental Data
3. Reef Connectivity Dataset
4. Human Development Index
5. Marine Protected Areas
6. Reefs at Risk Dataset
7. Coupled Model Intercomparison Project (CMIP5) climate projections Dataset





# Overview of Data Pipeline

2897  
Image features  
extracted from  
VGG16 model



Tabular data from  
Datasets 2-6  
(47 features)



Combined into a  
comprehensive coral  
reef feature set  
(11867 rows)

	date	latitude_degrees	longitude_degrees	depth_m	turbidity	cyclone_frequency
0	2002-08-30 00:00:00	-16.5348	-151.7299	3.0	0.0261	52.33
1	2002-09-03 00:00:00	-16.5353	-151.7284	1.8	0.0261	52.33
2	2002-09-05 00:00:00	-16.5427	-151.7408	1.5	0.0278	52.33
3	2002-09-08 00:00:00	-16.5397	-151.7887	0.8	0.0258	47.68
4	2002-09-09 00:00:00	-16.5526	-151.7326	0.8	0.0278	52.33

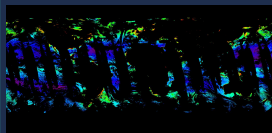
Feature Engineering

- Data Cleaning
  - **10 features**
- Standardization
- Feature Encoding

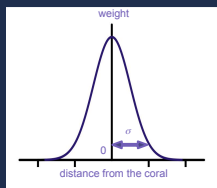


# Data Pipeline

MODIS  
Image Data



Gaussian Distance  
Weighting with  
 $\sigma = 1.0$



Weighted  
Images



4320 x 2160 x 3

- Gaussian function applies a **bell curve** to assign weights
- **Closer points** = more weight
- **Farther points** = reduced influence

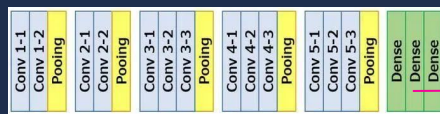
**VGG16** is a pre-trained CNN model used for feature extraction from images

Resize  
Images



224 x 224 x 3

Transfer Learning  
(CNN-VGG16)

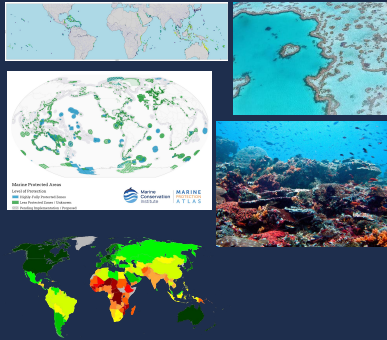


Extract  
features from  
second-to-last  
CNN layer



# Data Pipeline

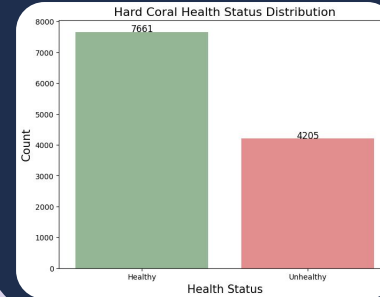
## Datasets 2-6



## Select 47 Relevant Features

- Geospatial and Temporal Data
- Environmental Conditions
- Ecological and Geographic Context
- Human Impact Metrics

## Categorize Target Variable: Hard Coral Class



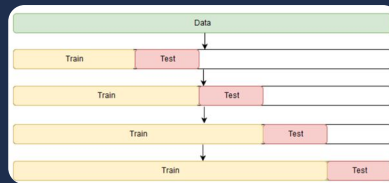
- Target variable derived from Hard Coral Percent Cover feature
- Percent cover  $> 20\%$  = Healthy reefs, dominant presence of hard corals
- Percent cover  $\leq 20\%$  = Unhealthy reefs, significant coral loss

# Modeling Overview

## Machine Learning Models

1. Logistic Regression
2. Random Forest Classifier
3. Extreme Gradient Boosting (XGBoost)
4. Support Vector Machine (SVM)
5. K-Nearest Neighbors (KNN)
6. Neural Network - Multi-Layer Perceptron (MLP) Classifier

Minimized AUC-ROC score using 5-fold Time Series Cross-validation

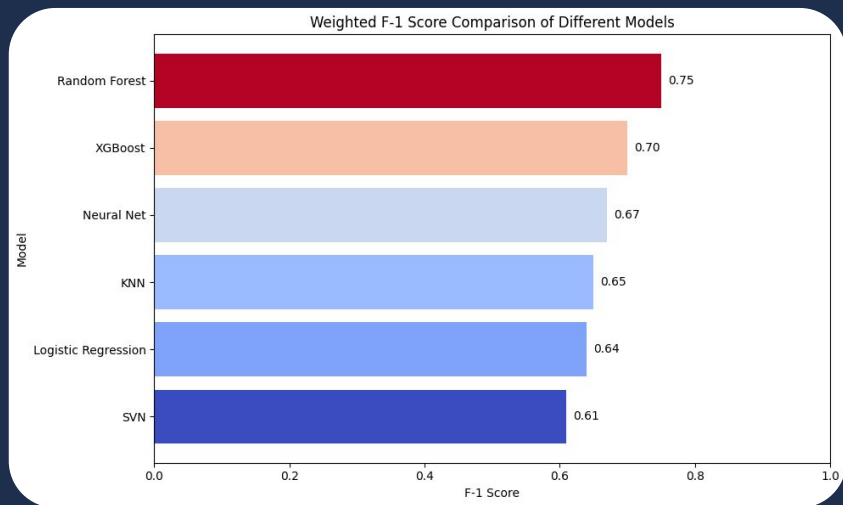


Best Model for Coral Health Prediction



# Model Performance (Hard Coral Class)

## Weighted F1 Scores



Evaluated multiple models and found **Random Forest** performed best for binary classification on test data metrics.

## RF Model Results: balanced training classes

Metric	Score
Weighted F-1	0.75
Accuracy	0.75
AUC-ROC	0.82

## RF Model Hyperparameters

- Max Depth: None
- Min Samples per Leaf: 1
- Min Samples per Split: 2
- Number of Estimators (Trees): 500

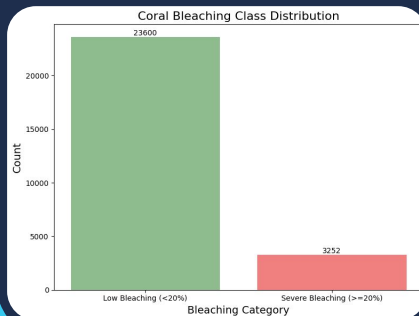
# Time Series Forecasting

Tabular data from  
Datasets 2-6

**(10 features)**



Categorize  
Target Variable:  
Percent Bleaching



Model Training and  
Evaluation

- Logistic Regression (baseline)
- Random Forest



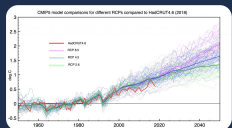
Best  
Model

- Target variable derived from Percent Bleaching feature
- Percent Bleaching  $\leq 20\%$  = Low Bleaching, healthy corals (retain color and algae)
- Percent Bleaching  $> 20\%$  = Severe Bleaching, stressed corals (lost color and vulnerable to diseases)

# Time Series Forecasting

## Data pipeline for forecasted values

CMIP5 climate projections



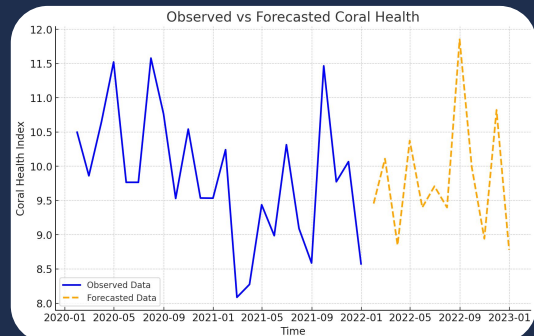
Select forecasted data for relevant features  
(2030, 2035, 2040)

Aggregate yearly for each reef

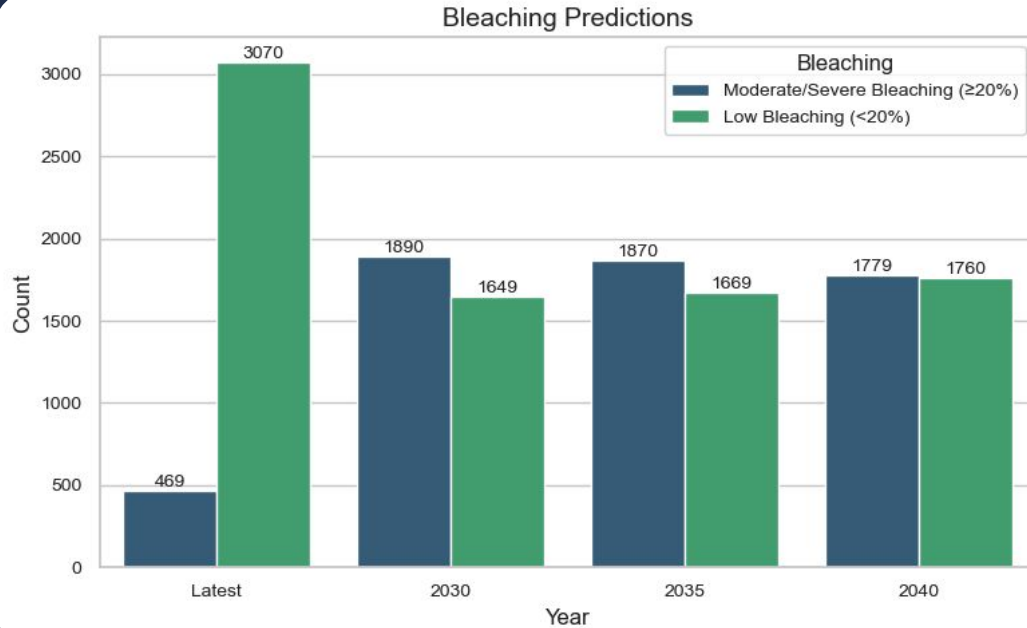


Best Model for Coral Health Prediction

## Time Series Forecasting (2030, 2035, 2040)



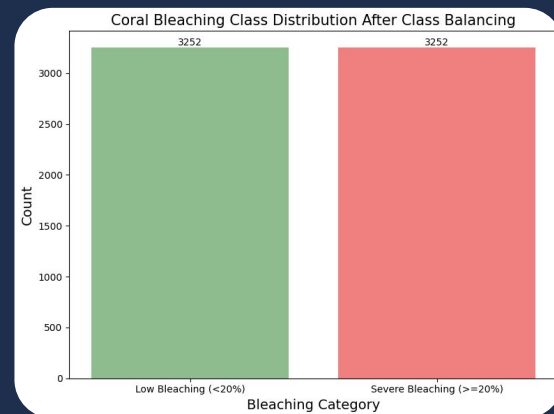
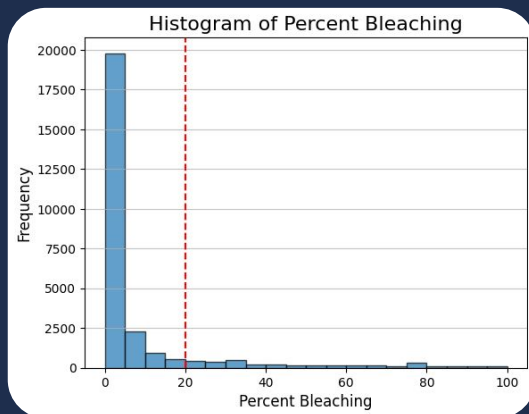
# Time Series Results



- Time series model using Random Forest
- 5, 10, 15 year look-ahead prediction
- Future predictions further out in time are more noisy and require additional analysis



# Model Performance (Bleaching)

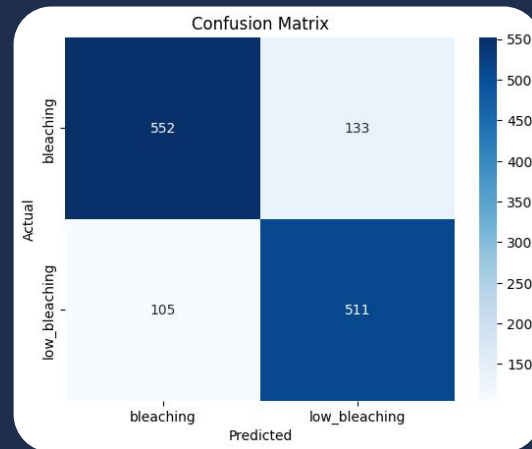


## RF Model Hyperparameters:

- Max Depth: 12
- Min Samples per Leaf: 4
- Min Samples per Split: 5
- Number of Estimators (Trees): 300

## RF Model Results:

Metric	Score
Weighted F-1	0.84
Accuracy	0.84
AUC-ROC	0.93



# Architecture

## Modeling

Colab + Google Cloud



- Quick and efficient
- Advanced GPU access for powerful training model development
- Free

## Deployment

Docker, .pkl File, Google Cloud, Flask API



- Easy
- Seamless integration
- Python - Flask
- Efficient
- Cost effective

## Website

Node.js, React, Vercel



- Industry standard: React and Node.js
- Easy hosting on Vercel
- Works well with docker and Google hosted API endpoints
- Free

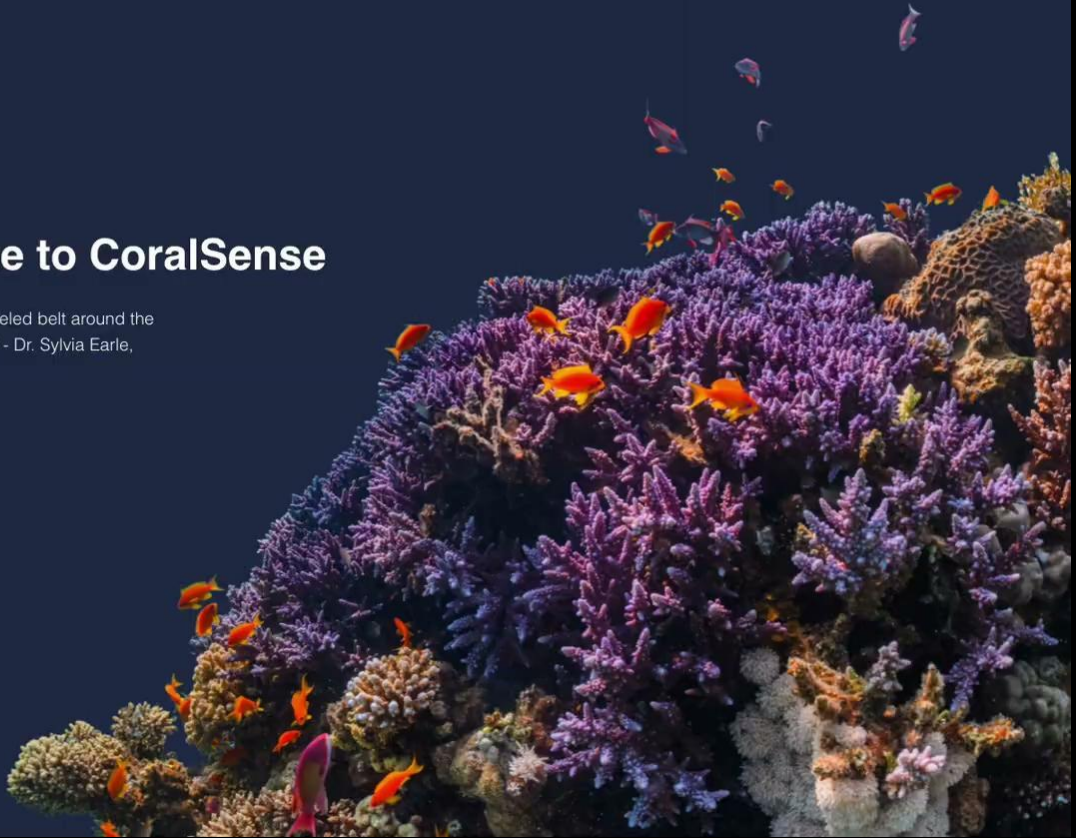
Architecture selected to optimize cost and MVP development speed

# What Can the Map Tool Do?

- Visualize Reef Locations
- Visualize Recent Coral Cover Records Globally
- Forecast the Future:
  - Visualize coral bleaching predictions through 2040 under an RCP 8.5, high carbon emissions scenario
- Enable Decision-Making:
  - Highlights high-priority reefs for immediate action and low-priority reefs for long-term strategies.
- Interactive Insights - users can:
  - Explore environmental factors affecting reef health.
  - Simulate the impact of changes in things like temperature or nutrient pollution on bleaching and hard coral cover

# Welcome to CoralSense

"Coral reefs are a jeweled belt around the middle of the planet." - Dr. Sylvia Earle, Oceanographer



# Next Steps



Integrate more feature projections into our prediction process



Develop educational and awareness promoting resources



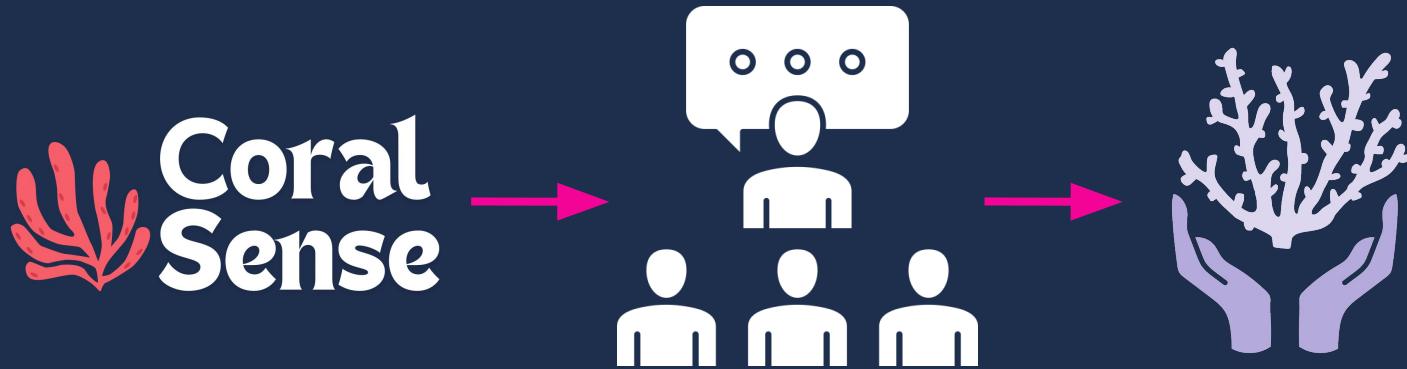
Collaborate with researchers collecting data on coral reefs



Promote advocacy to protect coral reefs

# The Quest for Change

- CoralSense equips researchers, conservationists, and policymakers with the tools they need to make informed decisions.
- Together, they can protect coral reefs, support biodiversity, and safeguard the livelihoods of millions of people worldwide.



# Our Mission

“Our mission is to use machine learning to monitor and model global coral reef health, providing accessible tools and insights to protect these vital ecosystems. By empowering communities and policymakers with actionable data, we aim to ensure a sustainable future for coral reefs and our oceans.”



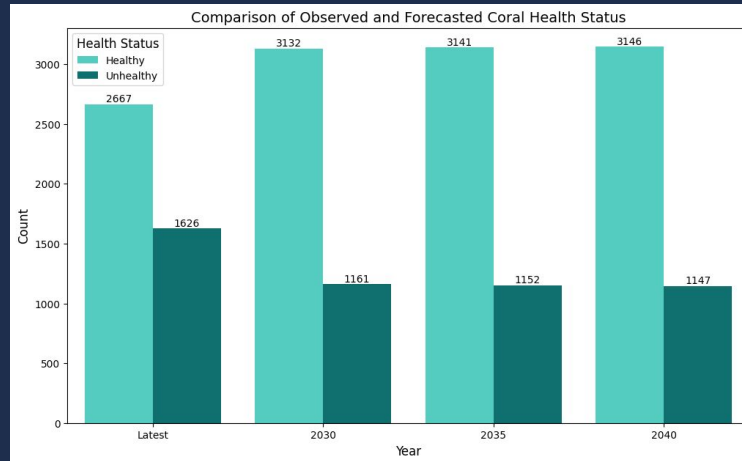


#ThriveAgain!



# Model Performance

Time Series

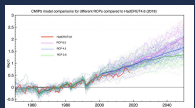


- Time series model using Random Forest
- 5, 10, 15 year look-ahead prediction
- Future predictions further out in time are more noisy and require additional analysis

# Time Series Forecasting

## Data pipeline for forecasted values

CMIP5 climate projections



Select forecasted data for relevant features  
(2030, 2035, 2040)

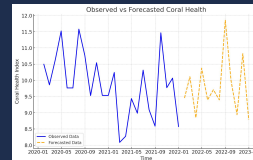
Aggregate yearly for each reef



Categorize Target Variable:  
Percent Bleaching

Select Best Model for  
Coral Health Prediction

Time Series Forecasting  
(2030, 2035, 2040)



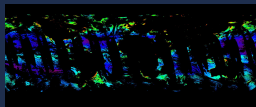
# The Magic of Prediction



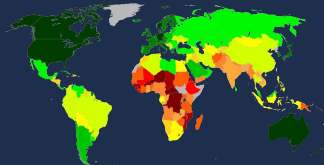
Reef connectivity



Global bleaching reef data



Modis image data



Human development index



Marine protected area



Reefs at risk

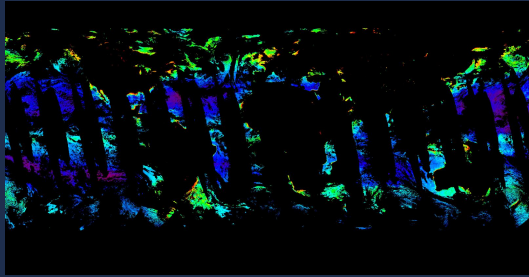
	date	latitude_degrees	longitude_degrees	depth_m	turbidity	cyclone_frequency
0	2002-08-30 00:00:00	-16.5348	-151.7299	3.0	0.0261	52.33
1	2002-09-03 00:00:00	-16.5353	-151.7284	1.8	0.0261	52.33
2	2002-09-05 00:00:00	-16.5427	-151.7408	1.5	0.0278	52.33
3	2002-09-08 00:00:00	-16.5397	-151.7887	0.8	0.0258	47.68
4	2002-09-09 00:00:00	-16.5526	-151.7326	0.8	0.0278	52.33

~ 12,000 Training records

Multiple large datasets engineered into a comprehensive coral reef feature set for model training

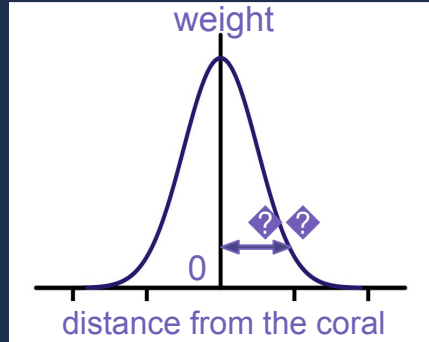
# Image Data - Preprocessing

Full Image Data



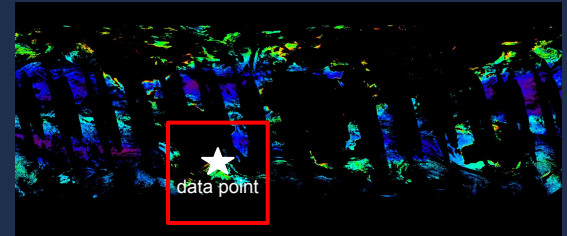
4320 x 2160 Images

Weighted Image Data



Gaussian Distance Weighting  
with  $\sigma = 1.0$

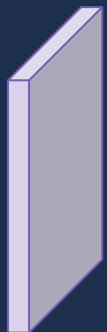
Segmented Image Data



**100 x 100 pixels** around data point

# Image Data - CNN (VGG16)

Weighted Images



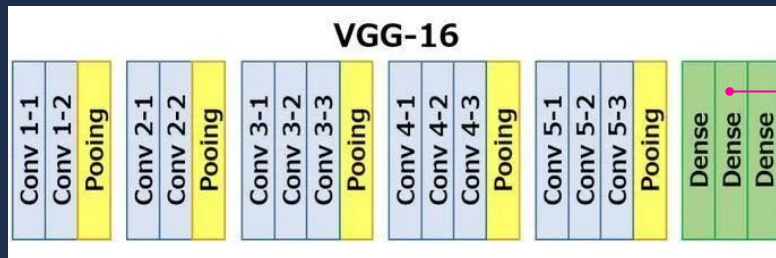
4320 x 2160 x 3

Resized Images

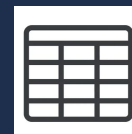


224 x 224 x 3

Transfer learning



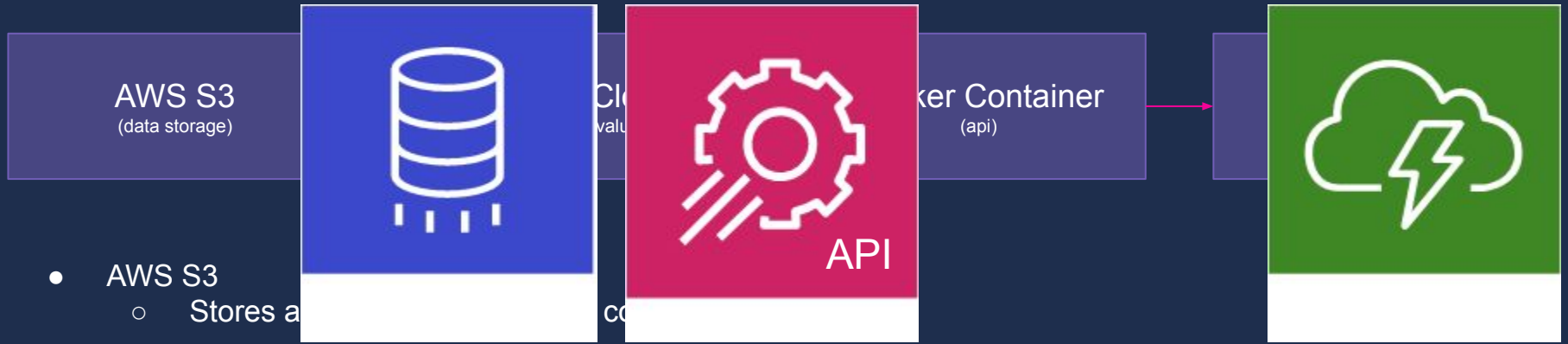
Extract features from the second-to-last CNN layer



Tabular Data

- Second-to-last layer retains high-level representations of input data
- Use both image and structured data to build models
- 11866 rows, 903 one-hot-encoded tabular data and 2897 image features

# Architecture



- AWS S3
  - Stores a
  - co
- Google Cloud
  - Handles model training, tuning, evaluation and version co
  - Google Kubernetes Engine
- Interactive Map
  - Displays real-time predictions overlaid on geographical



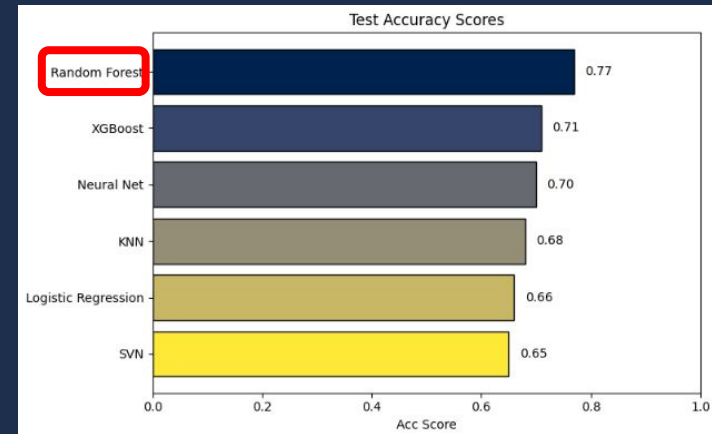
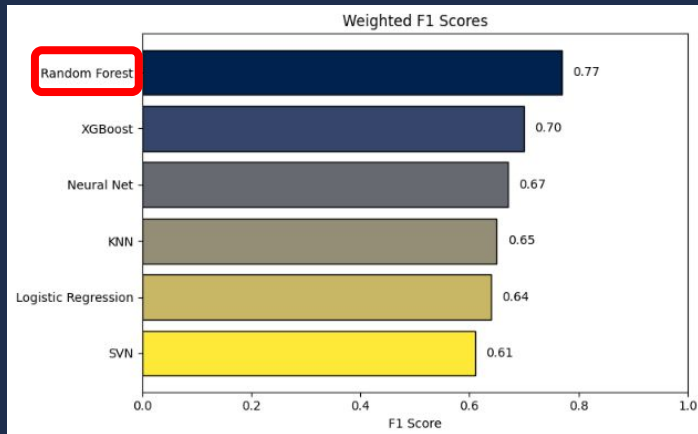
# Evaluation Results

2-class model yielded the best results

Model	2 Classes		3 Classes	
	Accuracy	Weighted F1-score	Accuracy	Weighted F1-score
Logistic Regression (baseline)	0.66	0.64	0.45	0.45
Random Forest	0.77	0.77	0.62	0.62
XGBoost	0.71	0.70	0.50	0.50
SVM	0.65	0.61	0.47	0.47
KNN	0.68	0.65	0.47	0.46
NN (MLPClassifier)	0.70	0.67	0.44	0.43

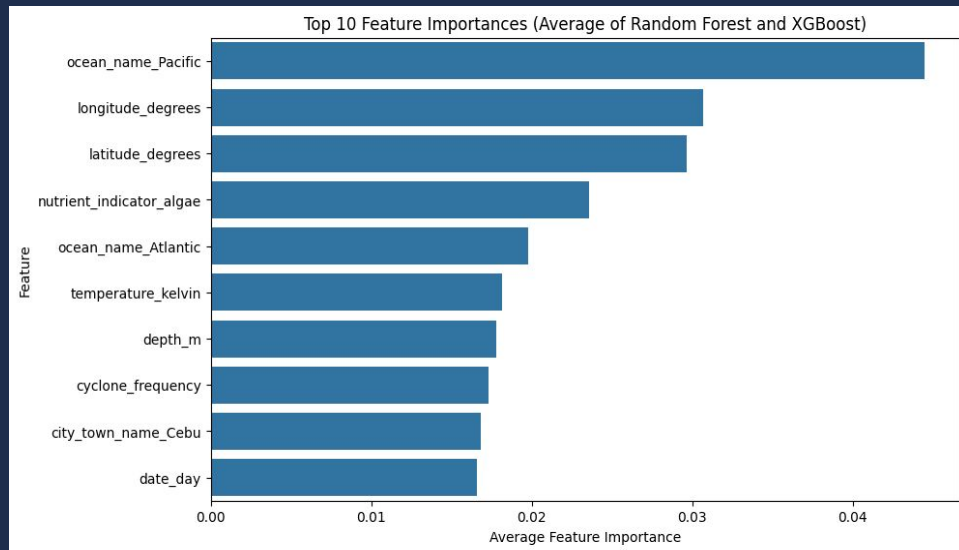
# Classification Model Performance

Evaluated multiple models and determined that a Random Forest model for binary classification performed best on test data in accuracy and weighted F1-score.





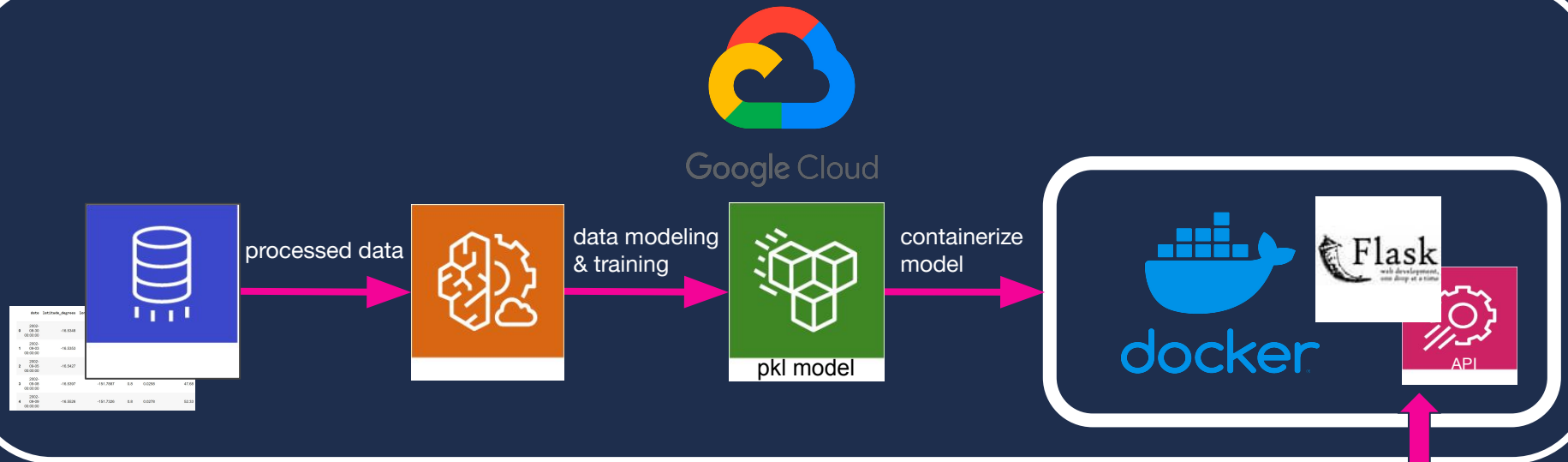
# Feature Importance



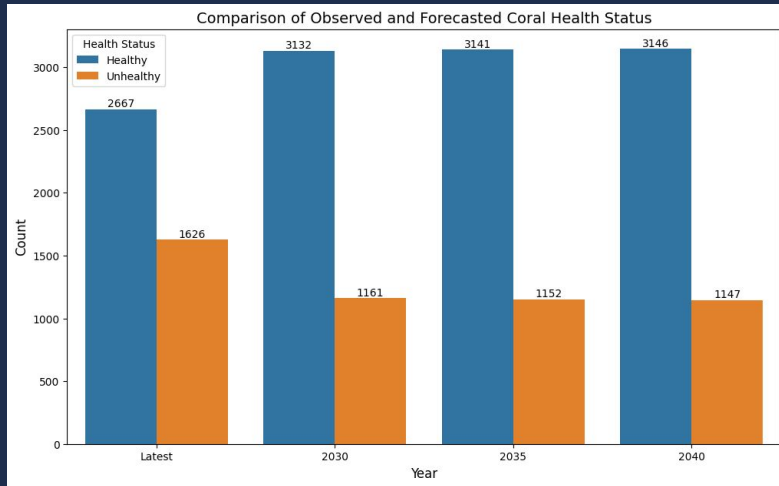
Significant in assessing coral health:

- Specific location data
- Environmental dynamics
- Habitat characteristics
- Seasonal fluctuations

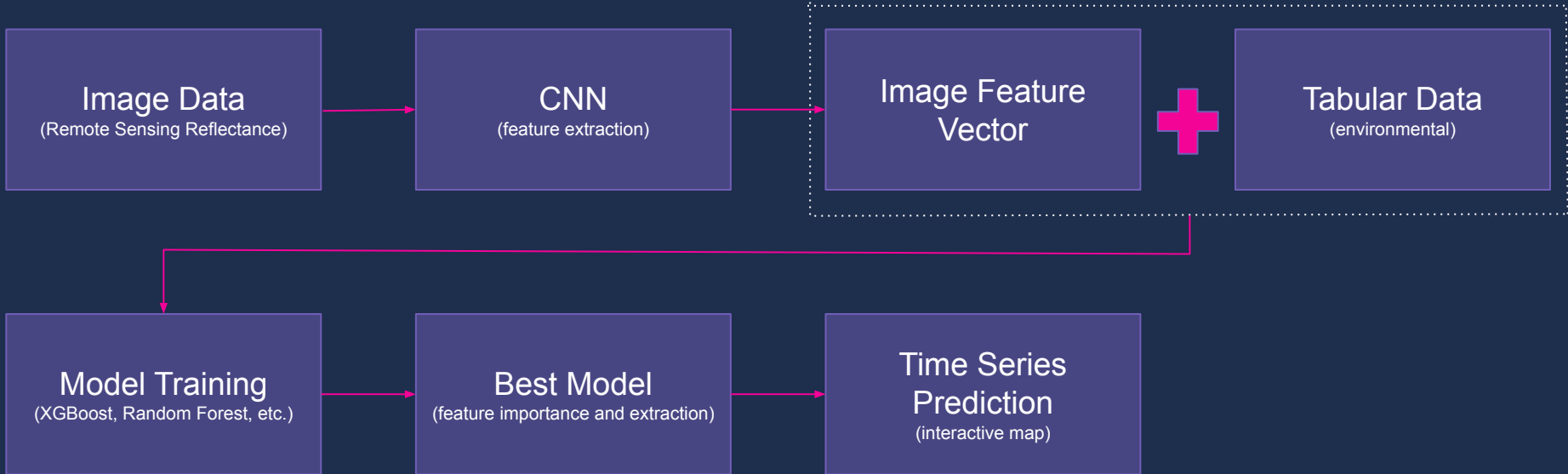
# Architecture



# Time Series Forecasting



# Modeling Overview



# Datasets

Dataset	Period	Records	Features	Source	Comments
Global Bleaching Environmental Data	1980 - 2020	41361	62	Biological & Chemical Oceanography Data Management Office	Main dataset Hard Coral Water temperature, reef location
Reef Connectivity		54596	178	Hawthorne et al 2018	reef connectivity is a measure of how well reefs can exchange organisms with other reefs, has been shown to help resilience
MODIS Aqua Level-3 Remote Sensing Reflectance (Rrs) Image Dataset	2002 - 2020	6749	2160 x 4320 RGB	NASA	high-resolution data on ocean biochemical properties
Human Development Index	1990 - 2022	6400+	36	United Nations Cinner et. al 2016	HDI is highly correlated with reef health
Marine Protected Areas	1984 - 2024	305012	31	Protected Planet	Cinner et. al 2016 showed that well-enforced marine protected areas correlate with better reef health
Reefs at Risk	2011		3	Reefs at Risk	Overfishing, marine pollution, and coastal development ratings

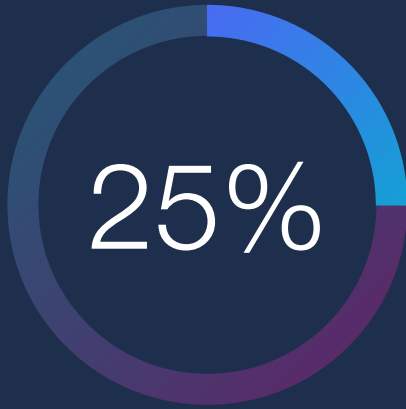
# Usability Study

Changes made due to usability study:

- Accessible color option
- Additional visuals on site
- Clearer instructions
- Addition of “Get Involved” page to promote education and advocacy

# SYMBOLS





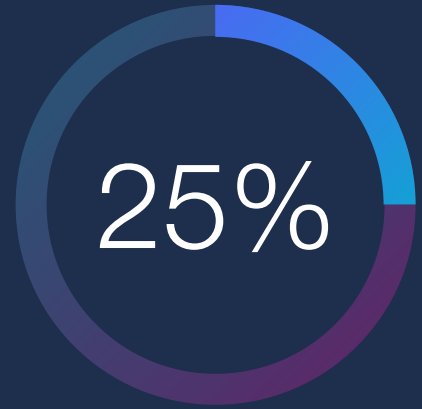
Of all marine life is in  
reefs



People depend on  
reefs



Coastline protected  
by reefs



Of fish caught in  
developing countries  
is from reefs



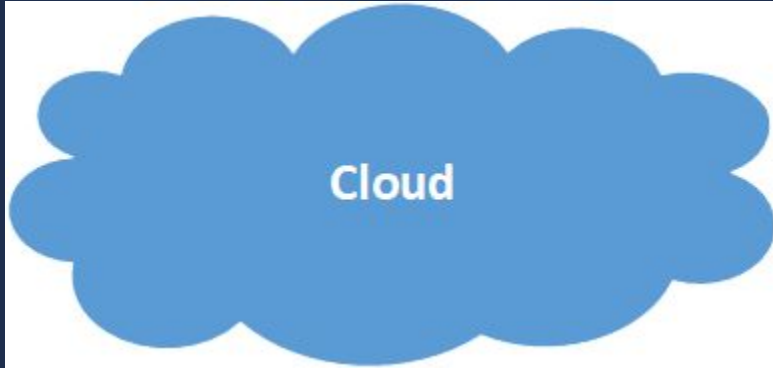
# SYMBOLS

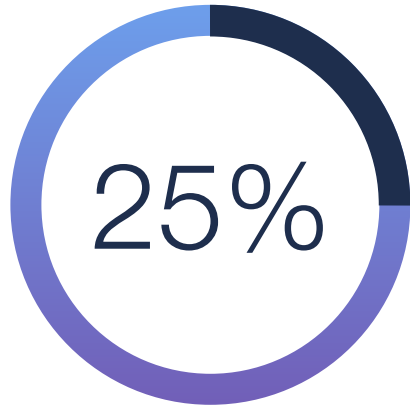


Google Cloud



docker





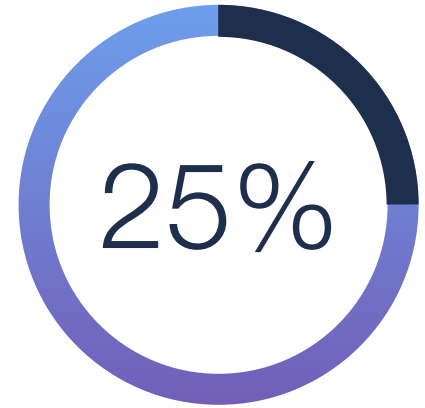
Of all marine life is in reefs



People depend on reefs



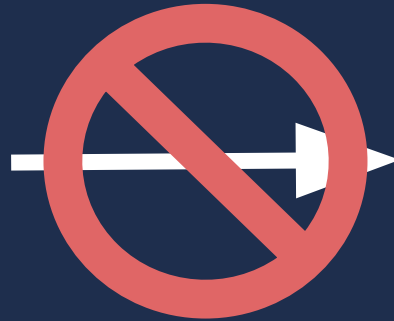
Coastline protected by reefs



Of fish caught in developing countries are from reefs

# The Call for Action

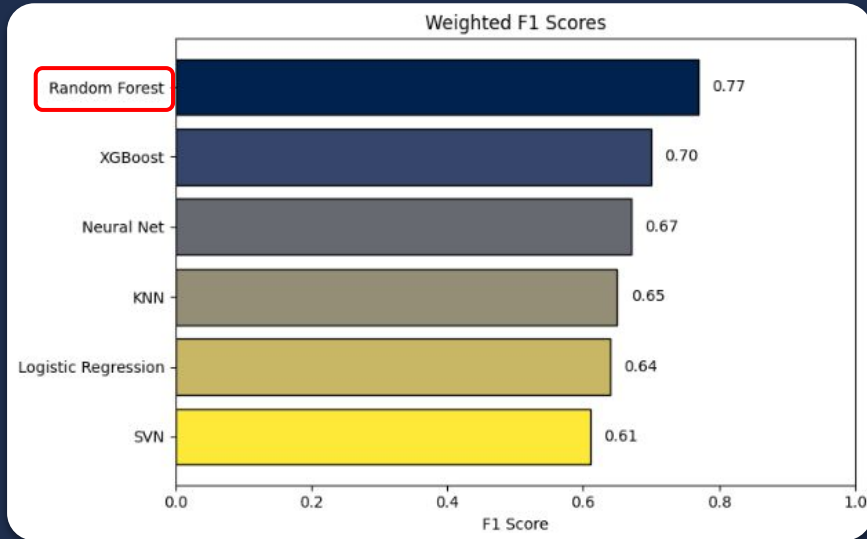
- The world calling for heroes to rise.
- Researchers, conservationists, and policymakers are seeking ways to protect what remains.
- But their tools are outdated, reactive instead of proactive, and cannot predict where to act before it is too late.



# Model Performance (Classification)

Evaluated multiple models and found Random Forest performed best for binary classification on test data metrics

Weighted F1 Score



Accuracy Score

