Developing a Recommender System for Shark Presence along East Coast Beaches

UNC CHARLOTTE

Department of Computer Science
College of Computing and Informatics

EST. 1851
CATAWBA
Salisbury, North Carolina

SAS Deep Learning Symposium

Lavanya Loganarayanan, MS, VP Retail Analytics Citi
Pamela Thompson, PhD, Associate Professor

Summer 2015

UNC CHARLOTTE

What caused so many attacks?

What caused so many attacks?

Humans are not deliberately being targeted.

"There's something going on there, there's no doubt about that . . . it's a PERFECT STORM of environmental and biological variables as well as human activity".

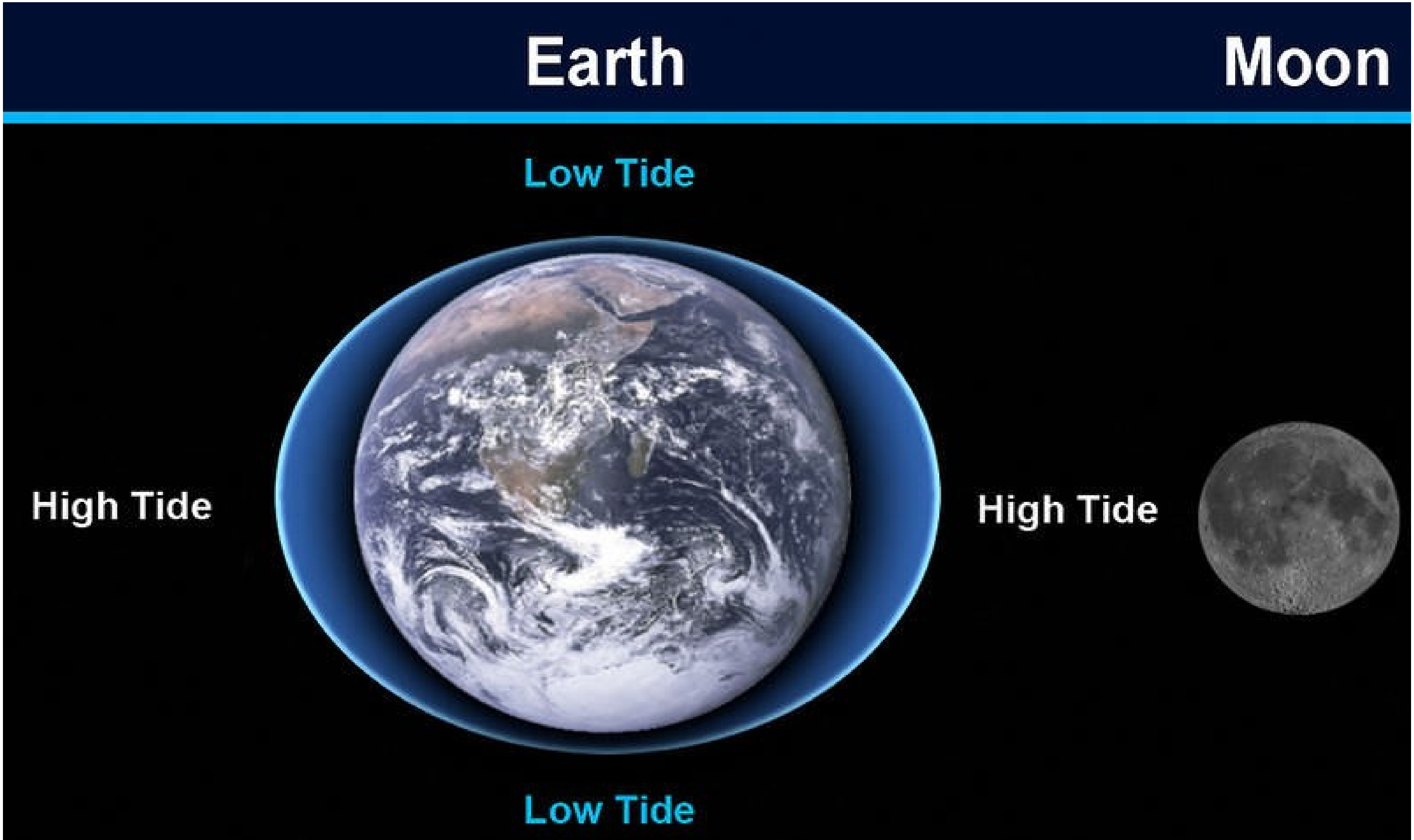
George H. Burgess, retired Director of the International Shark Attack File at the University of Florida's Florida Museum of Natural History

Theories

Population – Increased Human Activity

Theories

Moon Phases

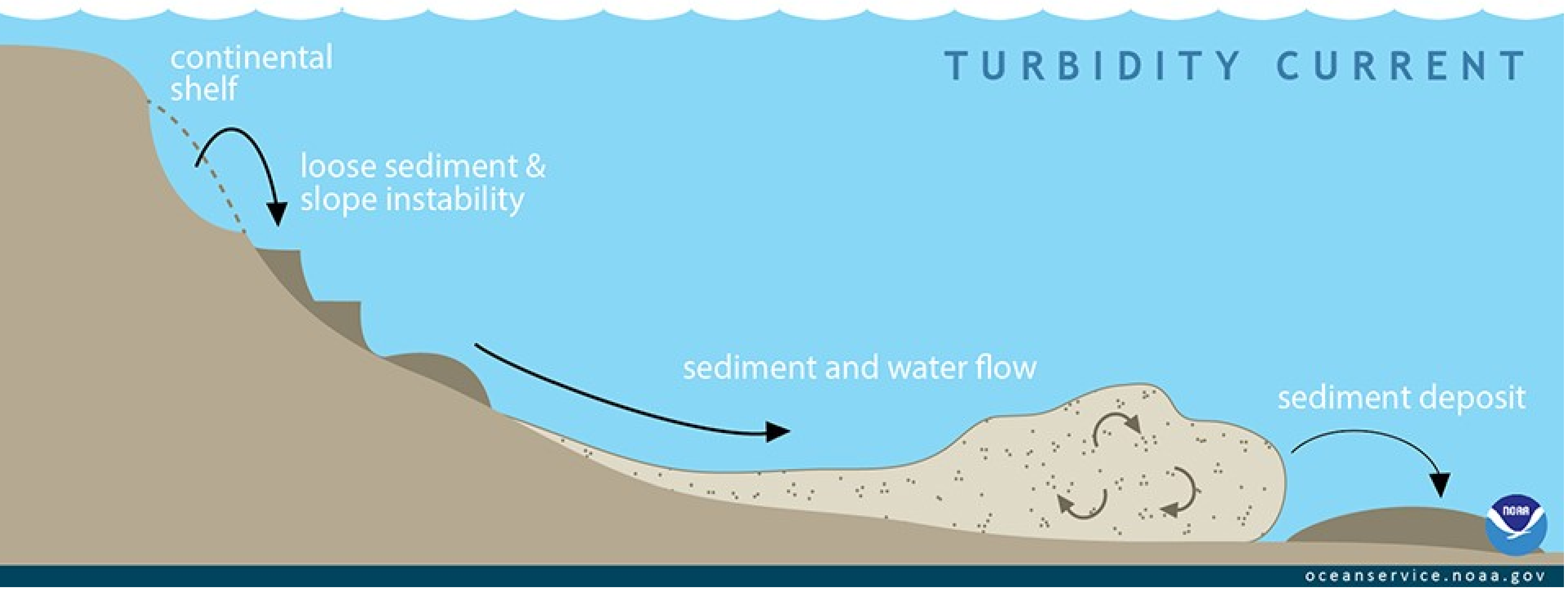


Theories

Weather Changes, Global Warming

Theories

Water Conditions



UNC CHARLOTTE

Theories

Shark Preservation Efforts

Theories

Plentiful Food

Research Statement of Purpose

. . to improve the understanding of the presence of sharks in near shore waters during tourist seasons in middle Atlantic and south eastern coastal waters, specifically North Carolina and South Carolina.

Benefits of this Study

Increased safety of humans participating in water activities

Increased understanding for greater protection for sharks

Economic, human safety, and environmental benefits

The model can be extended to other geographic areas

Benefits of this Study

Recommender System – much like the weather channel

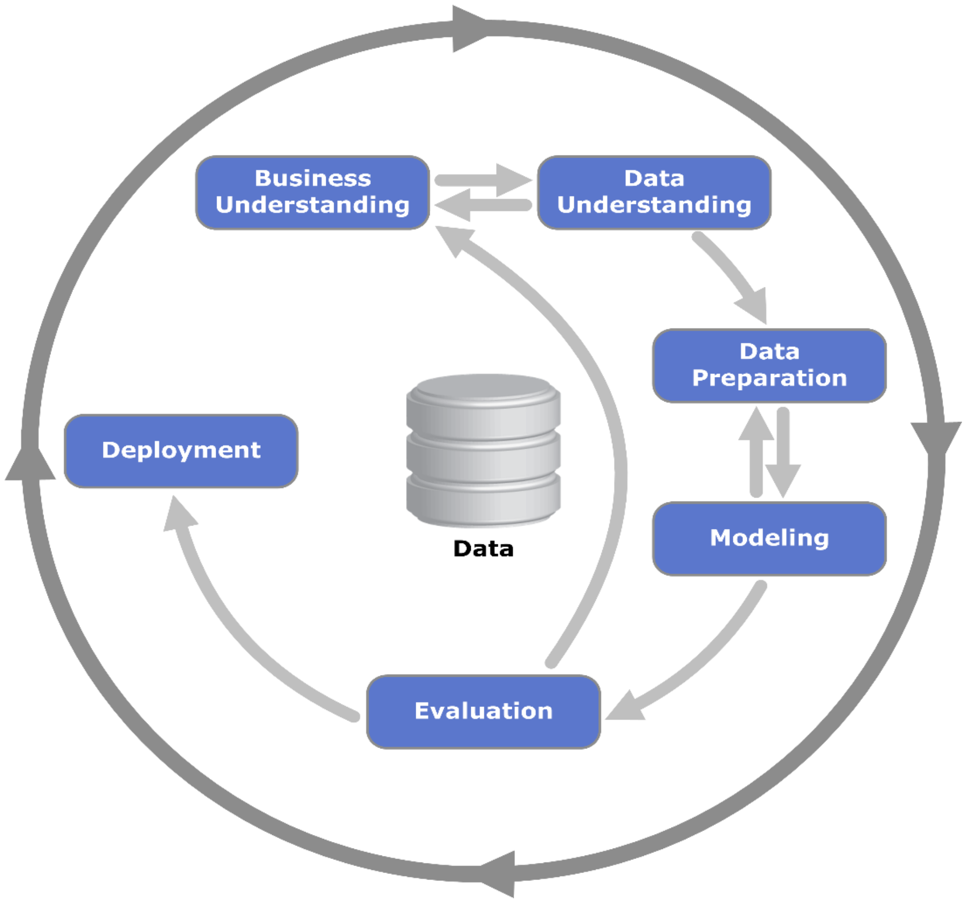
- Educates the public about the importance of the species
- Predicts Shark Presence
- Provides safety measures

Benefits of this Study

Promote interest in STEM and Data Science

The process: CRISP-DM

Cross Industry Standard Process – Data Mining



Domain Understanding

- Parts of North Carolina were abnormally dry
- The salinity of ocean water close to shore is higher than usual
- Sea turtles, crabs, fish urtles may draw sharks to the North Carolina shores.
- Is it the warming ocean causing the sharks to follow prey that are migrating due to warmer temperatures perhaps caused by climate change?
- The annual migration of menhaden fish, a favorite shark food, appears linked to water temperature, which jumped 10 degrees in a week during the heat wave of 2015

- What is true?
- What are the other factors?
- Is there hidden knowledge that can be discovered?

Sources of Data: Attacks

Global Shark Attack File:

http://www.sharkattackfile.net/

UNC CHARLOTTE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Case Numb	Date	Year	Type	Country	Area	Location	Activity	Name	S	Ag	Injury	Fat (Y/	Tim	Species	Investigator or Source
293	1853.09.28	28-Sept-1853	1853	Unprovoked	USA	North Carolina	Morehead, Carteret County	Commercial Salvage Diving	Alfetto	M		No injury. Copper breastplate & harness bitten	N		White shark	C. Creswell, GSAF; Washington Post, 10/7/1883, p.2
301	1853.00.00.c	Sep or Oct-1853	1853	Unprovoked	USA	North Carolina	Morehead, Carteret County	Hard hat diving	Mark Dare	M		No injury, copper breastplate punctured	N		White shark	Fort Wayne Gazette, 1/24/1897
384	2015.07.04.a	04-Jul-2015	2015	Unprovoked	USA	North Carolina	Off Surf City, Pender County		a marine	M	32	Lacerations to right hand & forearm	N	Evening		C. Creswell, GSAF
405	2015.07.01	01-Jul-2015	2015	Unprovoked	USA	North Carolina	Ocracoke, Lifeguard Beach, National Park Service, Hyde County	Swimming	Andrew Costello	M	68	Injuries to torso, hip, lower leg & hands	N	12h10	6' to 7' shark	C. Creswell, GSAF
412	2015.06.27.b	27-Jun-2015	2015	Unprovoked	USA	North Carolina	Rodanthe, Dare County	Swimming	John Cole	M	18	Injuries to right calf, buttock and both hands	N	16h00	Bull shark	C. Creswell, WRAL, 6/27/2015
421	2015.06.25	25-Jun-2015	2015	Unprovoked	USA	North Carolina	Avon, Hatteras Island, Outer Banks, Dare County	Body surfing?	Patrick Thornton	M	47	Multiple lacerations to back	N	11h41		C. Creswell, GSAF
436	2015.06.24.b	24-Jun-2015	2015	Unprovoked	USA	North Carolina	Surf City	Swimming	Brady Noyes	M	6	Minor injury to foot	N	12h25	Sandtiger shark	C. Creswell, GSAF
441	2015.06.14.b	14-Jun-2015	2015	Unprovoked	USA	North Carolina	Oak Island, Brunswick County	Wading	Hunter Treschel	M	16	Arm amputated below shoulder	N	17h51	Bull shark	C. Creswell, GSAF
446	2015.06.14.a	14-Jun-2015	2015	Unprovoked	USA	North Carolina	Oak Island, Brunswick County	Wading	Kiersten Yow	F	12	Left arm amputated at elbow & severe injury to leg	N	16h12	Bull shark	C. Creswell, GSAF
461	2015.06.11	11-Jun-2015	2015	Unprovoked	USA	North Carolina	Ocean Isle, Brunswick County	Boogie boarding	girl	F	13	Minor lacerations to foot	N	12h45	4' shark	C. Creswell, GSAF
5541	2015.10.09.b	09-Oct-2015	2015	Unprovoked	USA	South Carolina	Shipyard Beach Club, Hilton Head Island, Beaufort County	Boogie boarding	Meti Kershner	F	9	Laceration to forearm	N	16h20		C. Creswell, GSAF
5596	2015.09.03	03-Sep-2015	2015	Unprovoked	USA	South Carolina	Myrtle Beach, Horry County		Chip Wagner	M		Right foot bitten	N	16h00	4' shark?	C. Creswell, GSAF
5672	2015.08.20	20-Aug-2015	2015	Unprovoked	USA	South Carolina	Murrells Inlet, Georgetown County	Surfing	Dylan Peyton	M	15	Injuries to left calf, arm and hand	N	12h30	4' shark	C. Creswell, GSAF
5698	2015.07.26.a	26-Jul-2015	2015	Invalid	USA	South Carolina	Edisto Beach, Colleton County	Floating	female	F	35	2' cut to dorsum of foot, 2 puncture wounds to sole		10h10	Thought to involve a 3' to 4' shark, but shark involvement not confirmed	C. Creswell, GSAF, ABC 11, 7/27/2015
5847	2015.06.30.a	30-Jun-2015	2015	Unprovoked	USA	South Carolina	Isle of Palms County Park, Isle of Palms, Charleston County	Playing in the water	Kysen Weakley	M	12	Shallow lacerations & puncture to lateral left leg	N	18h05	4' to 5' shark	C. Creswell, GSAF
5951	2015.06.26.a	26-Jun-2015	2015	Unprovoked	USA	South Carolina	South Beach, Hunting Island State Park, Beaufort County	Standing	Lance Donahue, Jr	M	43	Puncture wounds to foot	N	11h00	4' shark	C. Creswell, GSAF; WCNC, 6/26/2015
5978	2015.06.23	23-Jun-2015	2015	Unprovoked	USA	South Carolina	St. Helena Island, Beaufort County	Standing	male	M	9	Minor injury to calf	N		small shark	C. Creswell, GSAF; R. Lurye, Island Packet
6040	2015.05.15	15-May-2015	2015	Unprovoked	USA	South Carolina	Sullivan's Island		male	M	30	Laceration to foot	N	14h15	6' shark	News 2, 5/15/2015

Data: Weather

Temperature, Precipitation, Moving Average Precipitation, Wind Speed, Wind Direction

Daily readings from NOAA

Daily measures included in our research

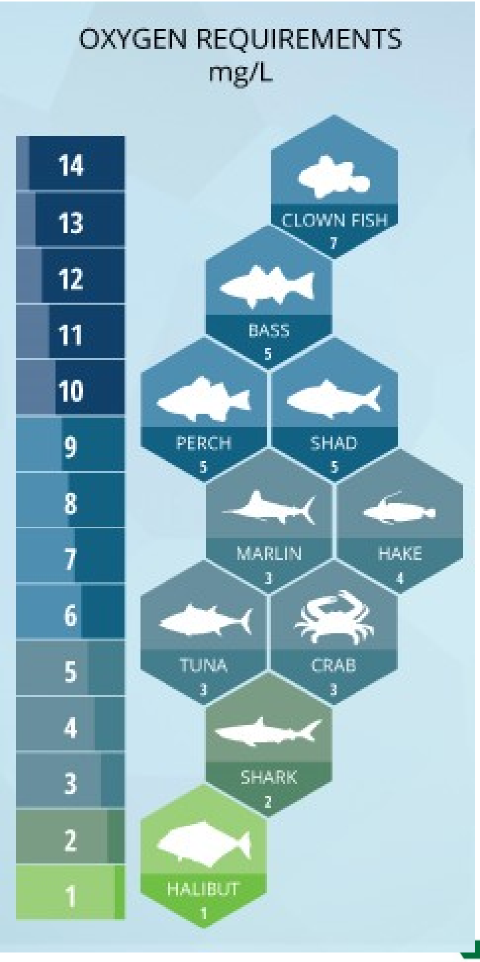
Days of attacks

Days without attacks included

Summer months

Data: Water

- Salinity, Turbidity, Oxygen, Sea Water Temperature from East Cribbing Station in NC
- Oxygen requirements
- Low Oxygen Suspected as cause of higher shark attacks due to low fish populations



UNC CHARLOTTE

Data: Crabs

Sharks eat crabs.

Crabs have more frequent movement during full and new moon phases.

Data: Turtles

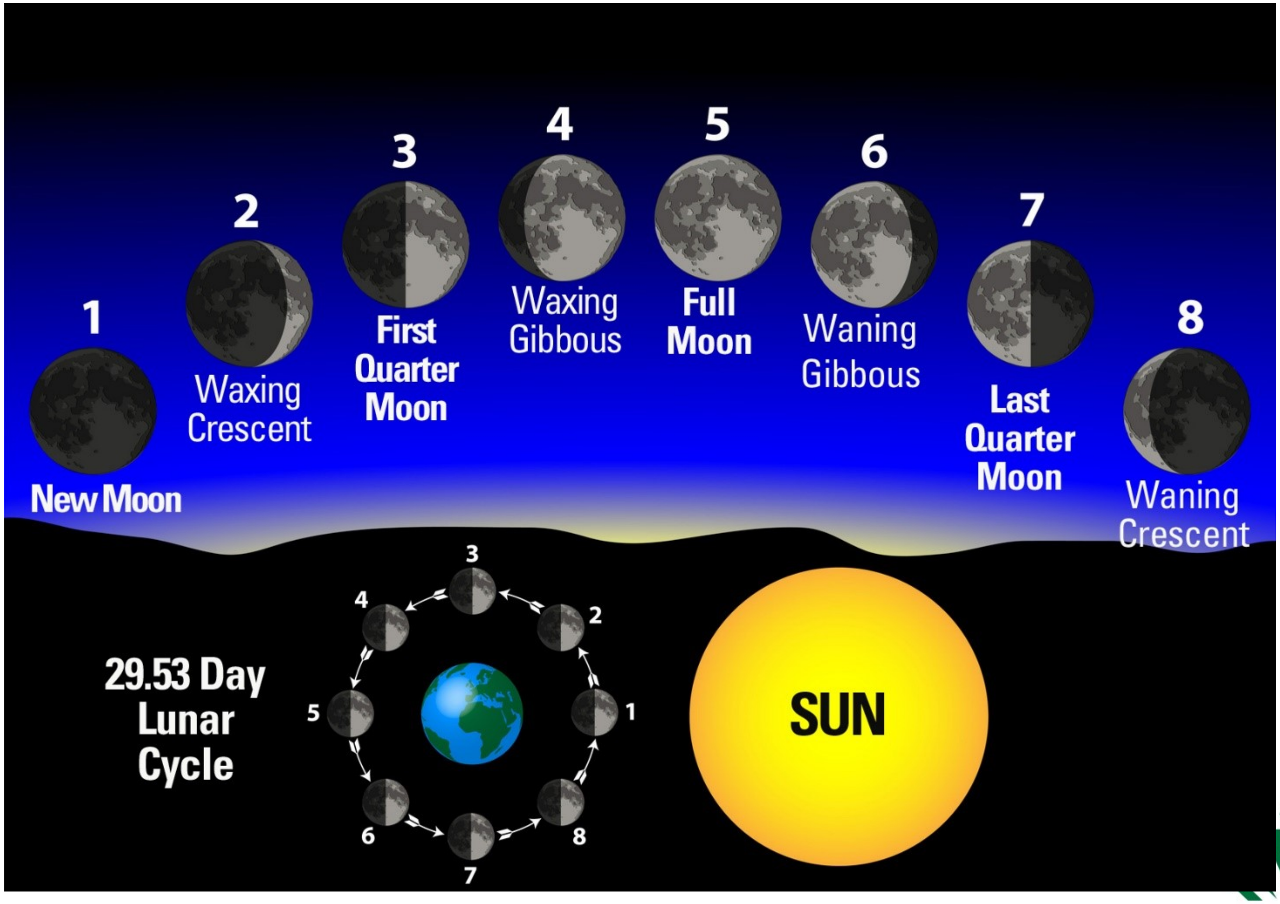
Sharks eat turtles - NC and SC have many nests
Turtles move to beaches for nesting and false crawl.

Privacy and security concerns with data

Turtle data from M. Pate and Dr. M. Godfrey
State Coordinators, SC-NC Wildlife Resources Comm.

Data: Moon Phases

Function for lunar calculations



Data Preparation

1) Numeric values: Discretized, Normalized for new attributes (3 bin)

2) Precipitation: 5 day moving average added – new feature

3) Moon phases: extended for Full, New Moon

4) Crab Data: Necessary imputations were made – weekend data not available

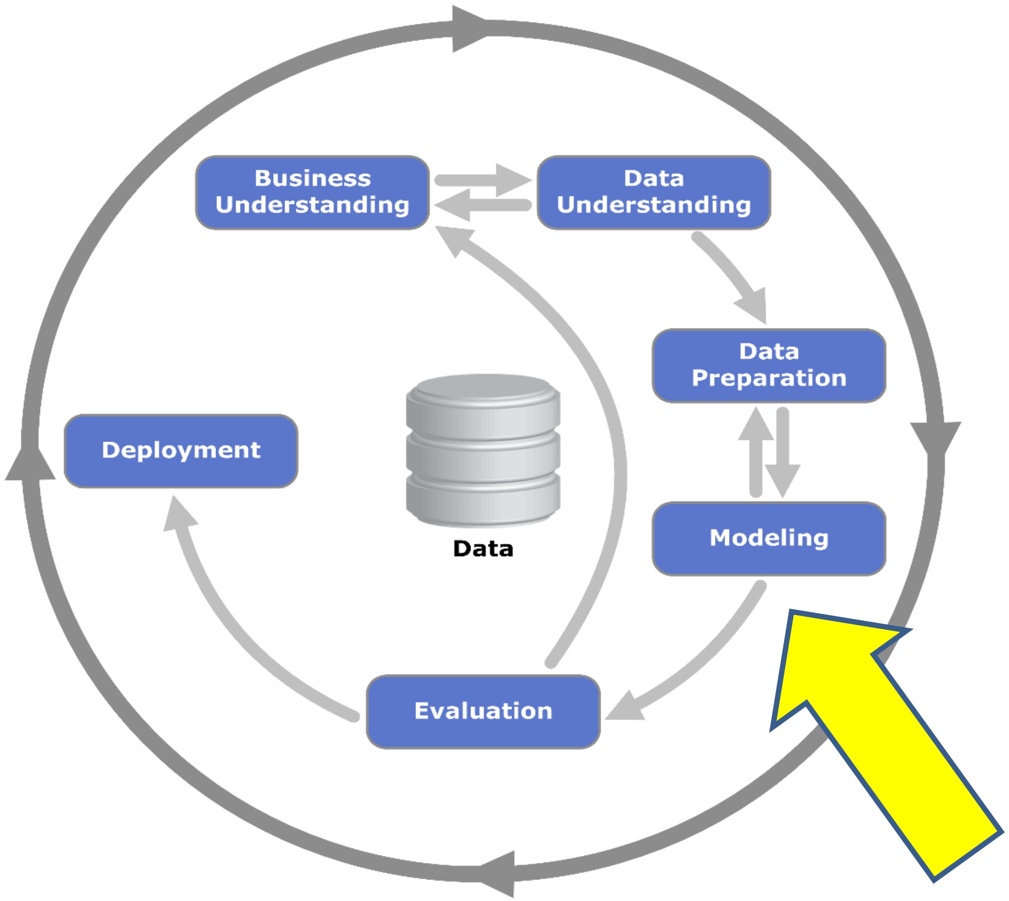
5) Class imbalance problem! (common with rare events and prediction)
Handled with *stratified* under sampling of Attack=No
Subset so that 1/3 of records remain with adequate representation for each date.

Oversampling of minority class may be better with rare events . . .

UNC CHARLOTTE

The process: CRISP-DM

Cross Industry Standard Process – Data Mining



EDA and Modeling: Clustering

Simple EM (Expected Maximization)

Cluster 2 included 46 attack = Y

We can learn from this cluster

Attribute	Cluster 0 (0.11)	Cluster 1 (0.06)	Cluster 2 (0.17)	Cluster 3 (0.65)
Turtle_Discretize				
Medium	6.1868	10.8648	20.8048	66.1436
Low	23.4523	3.2824	11.3272	106.9381
High	3.1305	6.7933	19.6859	11.3903
Very High	2.9729	1.1292	1.9987	4.8992
[total]	35.7424	22.0697	53.8166	189.3712
Attack				
No	31.965	15.8598	5.5238	170.6514
Yes	1.7775	4.21	46.2927	16.7198
[total]	33.7424	20.0697	51.8166	187.3712
MoonPhaseExtended				
Full	5.4434	[illegible]	16.705	45.4288
Third quarter	2.7519	[illegible]	[illegible]	25.0269
New	12.7777	[illegible]	14.3133	54.9404
First quarter	[illegible]	[illegible]	9.5607	21.9515
Waning gibbous	[illegible]	[illegible]	4.4876	15.8588
Waxing gibbous	3.7436	3.1043	5.8625	8.2897
Waning crescent	1.8039	2.3745	1.9528	9.8688
Waxing crescent	6.422	2.6731	2.8985	12.0064
[total]	39.7424	26.0697	57.8166	193.3712
Dissolved02_discretize				
Low	7.2264	3.2664	8.219	6.2882
Medium	17.1091	13.4138	41.5236	180.9535
High	10.4069	4.3895	3.074	1.1296
[total]	34.7424	21.0697	52.8166	188.3712
salinity_discretize				
High	1.5592	2.8721	19.4847	158.0839
Medium	31.9402	5.5168	31.3896	29.1534
Low	1.243	12.6808	1.9423	1.1339
[total]	34.7424	21.0697	52.8166	188.3712
turbidity_discretize				
Low	26.6803	15.8878	47.8739	184.558
High	2.0227	1.0415	1.0268	1.909
Medium	6.0394	4.1404	3.9159	1.9043
[total]	34.7424	21.0697	52.8166	188.3712
temperature_discretize				
High	24.126	9.5087	49.2901	186.0752
Medium	8.7728	10.504	2.5082	1.215
Low	1.8436	1.057	1.0183	1.0811
[total]	34.7424	21.0697	52.8166	188.3712

Modeling: Clustering

Simple EM (Expected Maximization)

Turtle – high

Moon Phase – full and new

Salinity is medium and high

Turbidity – low

Precipitation – low

Crab - high

```
pressure_discretize
  Medium                    23.7769   9.3064  31.2966 132.6201
  High                       5.9865   8.4276  19.3288  47.2571
  Low                        4.9791   3.3357   2.1912    8.494
  [total]                   34.7424  21.0697  52.8166 188.3712
windspeed_discretize
  Low                       22.4374    8.862   6.6265 129.0741
  Medium                     9.5352  11.1864  28.6779  57.6005
  High                       2.7698   1.0213  17.5122   1.6966
  [total]                   34.7424  21.0697  52.8166 188.3712
precipitationmva_discretize
  Low                       30.6367  17.9498    49.55 161.8635
  Medium                     1.2302   2.0038   2.2526  24.5134
  High                       2.8755   1.1162   1.0139   1.9944
  [total]                   34.7424  21.0697  52.8166 188.3712
Crab_Landings_Discretize
  High                       6.0063   9.2566  33.8044  41.9328
  Medium                    14.3684   2.2614   8.8114  73.5588
  Low                       14.3678   9.5517  10.2008  72.8797
  [total]                   34.7424  21.0697  52.8166 188.3712


Time taken to build model (full training data) : 5.74 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       25 (  9%)
1       15 (  5%)
2       48 ( 17%)
3      197 ( 69%)
```

UNC CHARLOTTE

The WILLIAM STATES LEE COLLEGE of ENGINEERING

Modeling: Association Rule Mining

Best rules:

5. MoonPhaseExtended=New DissolvedO2_discretize=Low 7 ==> Attack=Yes 7 conf:(1)

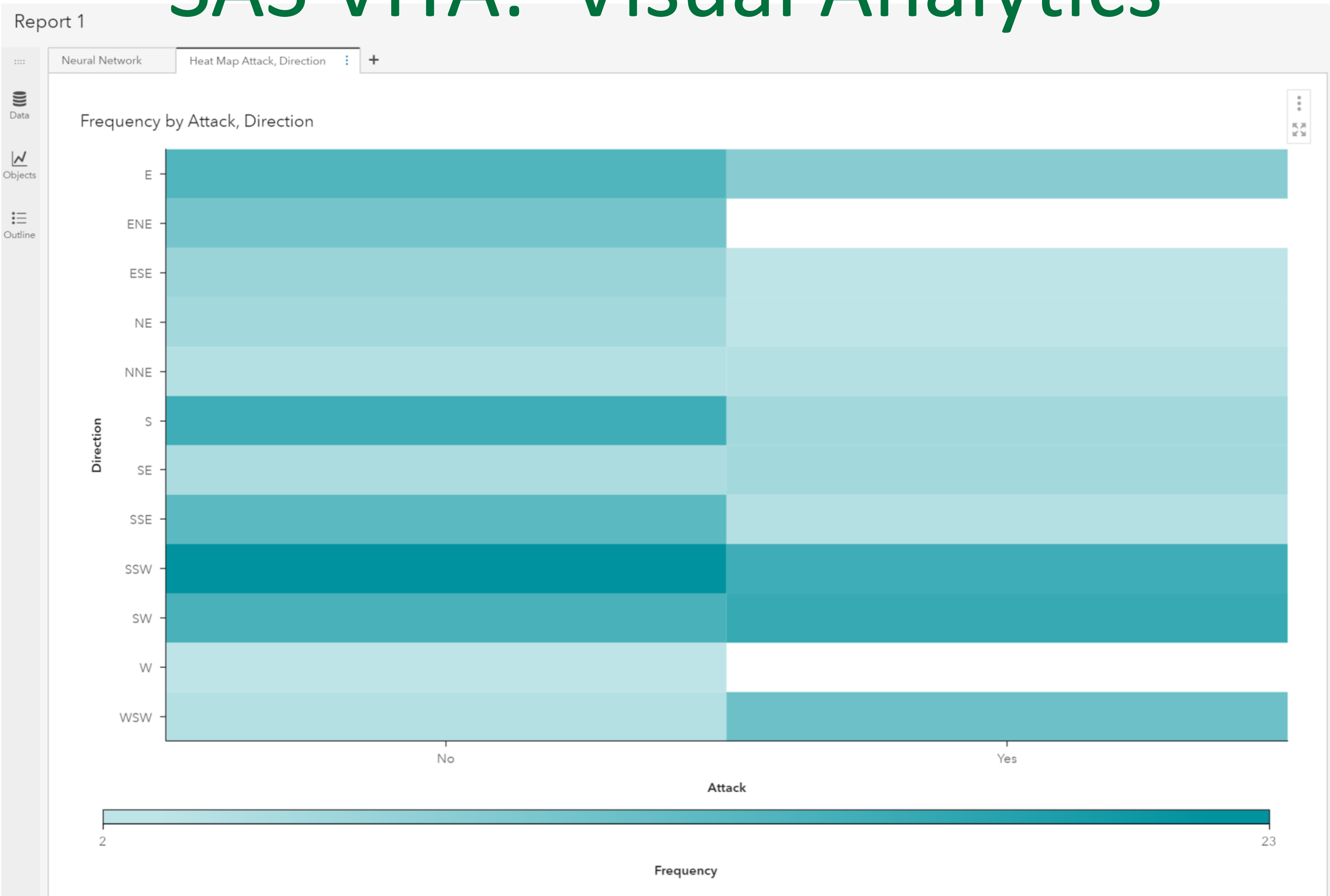
11. Turtle_Discretize=High MoonPhaseExtended=New 6 ==> Attack=Yes 6 conf:(1)

18. MoonPhaseExtended=New DissolvedO2_discretize=Low turbidity_discretize=Low 5 ==> Attack=Yes 5 conf:(1)

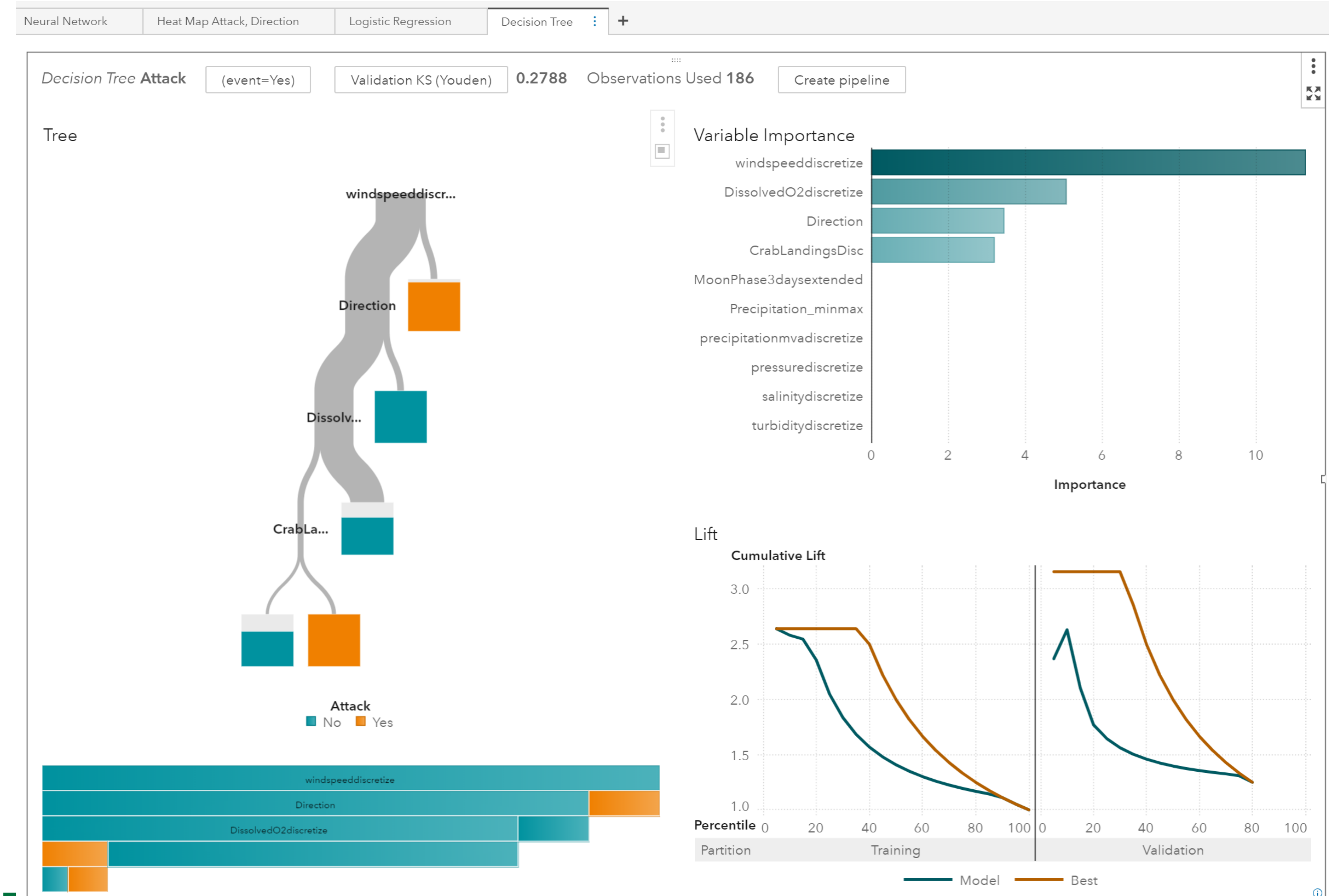
SAS® Viya™

Unified, powerful, adaptive, open

SAS VIYA: Visual Analytics



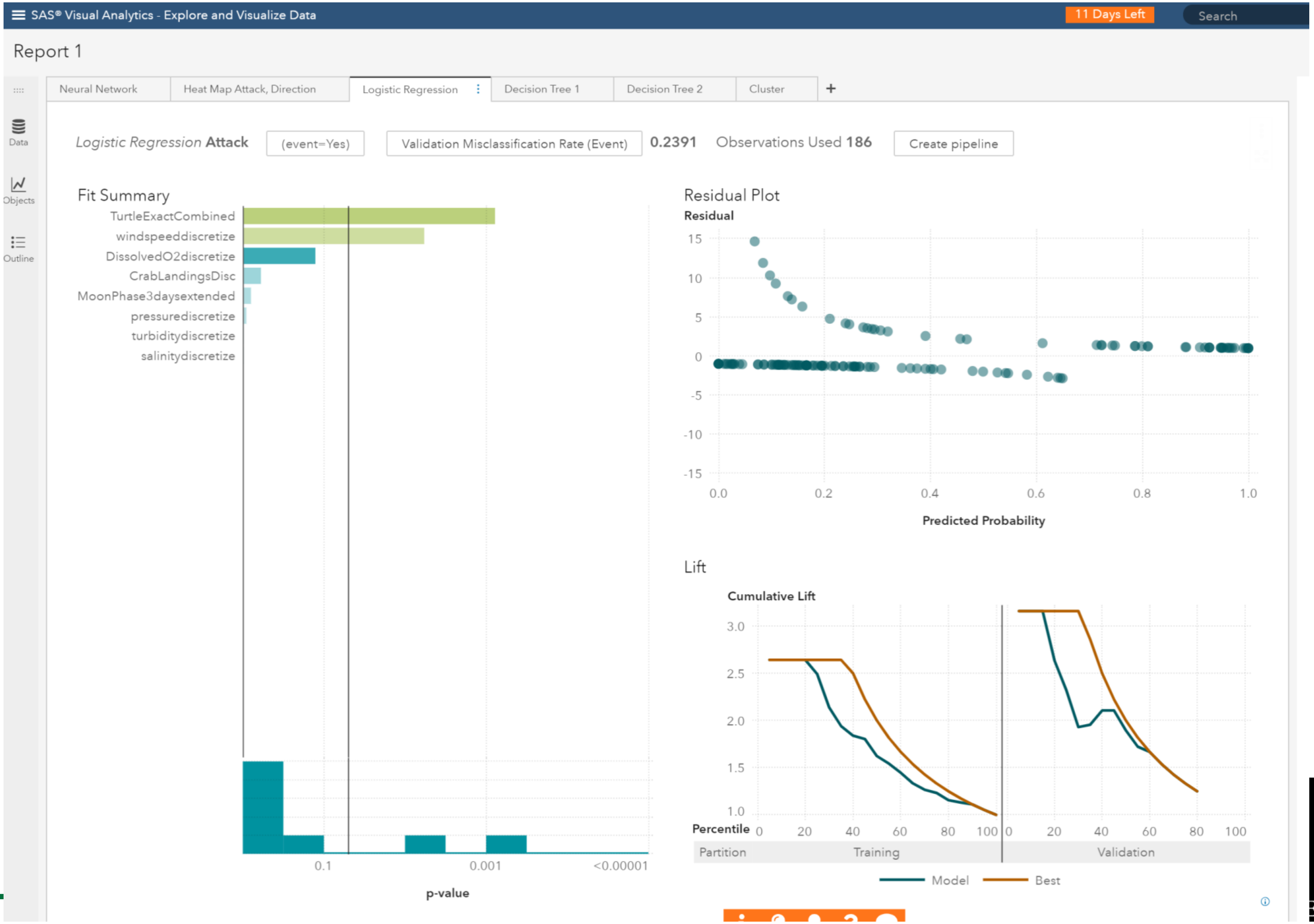
Decision Tree



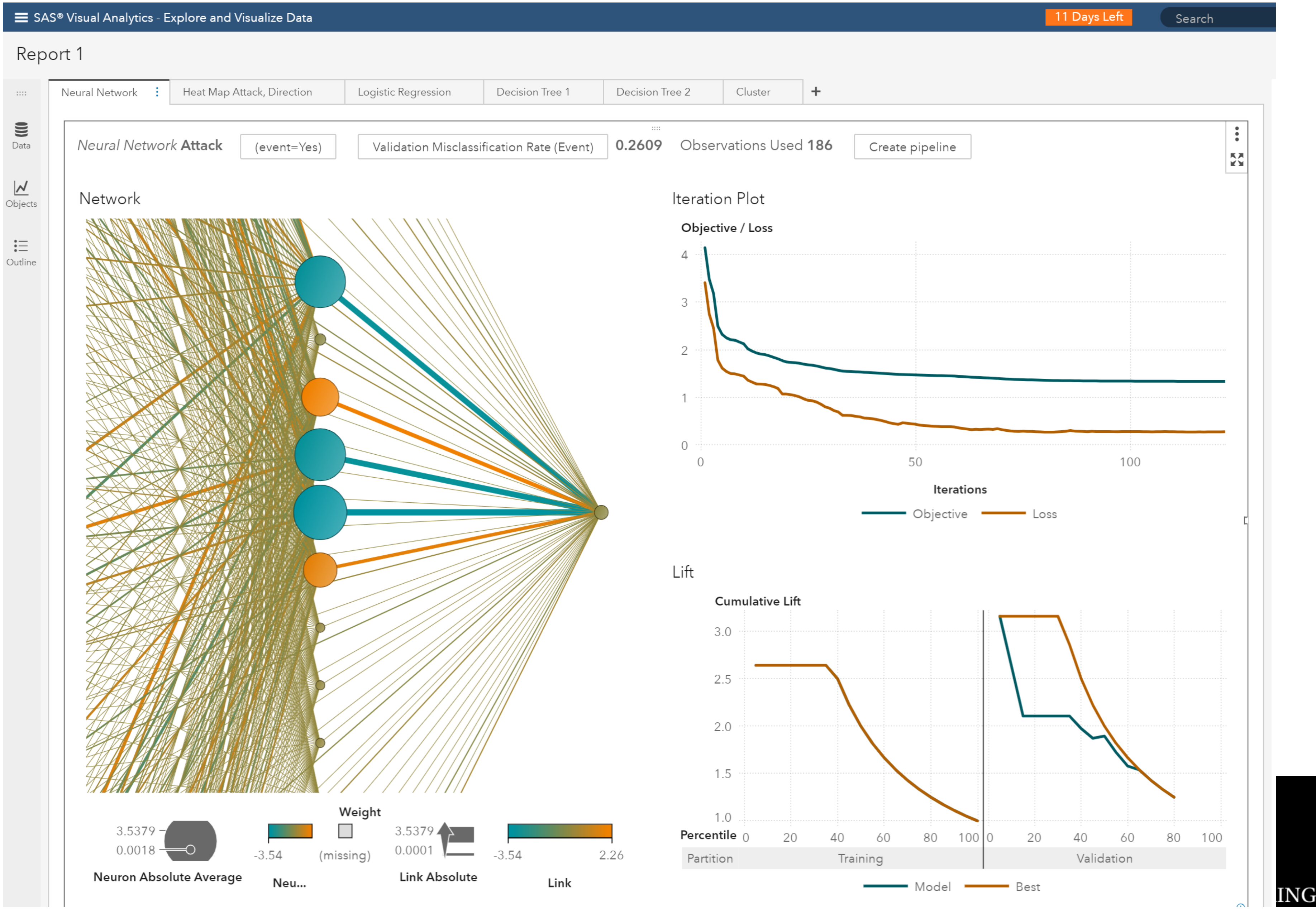
Clustering


Cluster Observations Used 186 Polylines 129
Wind...
Direc...
Moo...
Turbi...
0.8
0.0
NE
WSW
Third quarter
1.00
0.00
No
Yes
-0.25 0.00 0.25 0.50 0.75 1.00
First quarter
Full
New
Third quarter
Attack
WindSpeed_minmax
Direction
MoonPhase4daysextended
Cluster ID
1 2 3 4
Cluster ID
Attack
WindSpeed_minmax
Direction
MoonPhase4daysextended
Turbidity_minmax
1
0
Yes
No
SW
SSW
S
E
New
Full
1
0

Logistic Regression

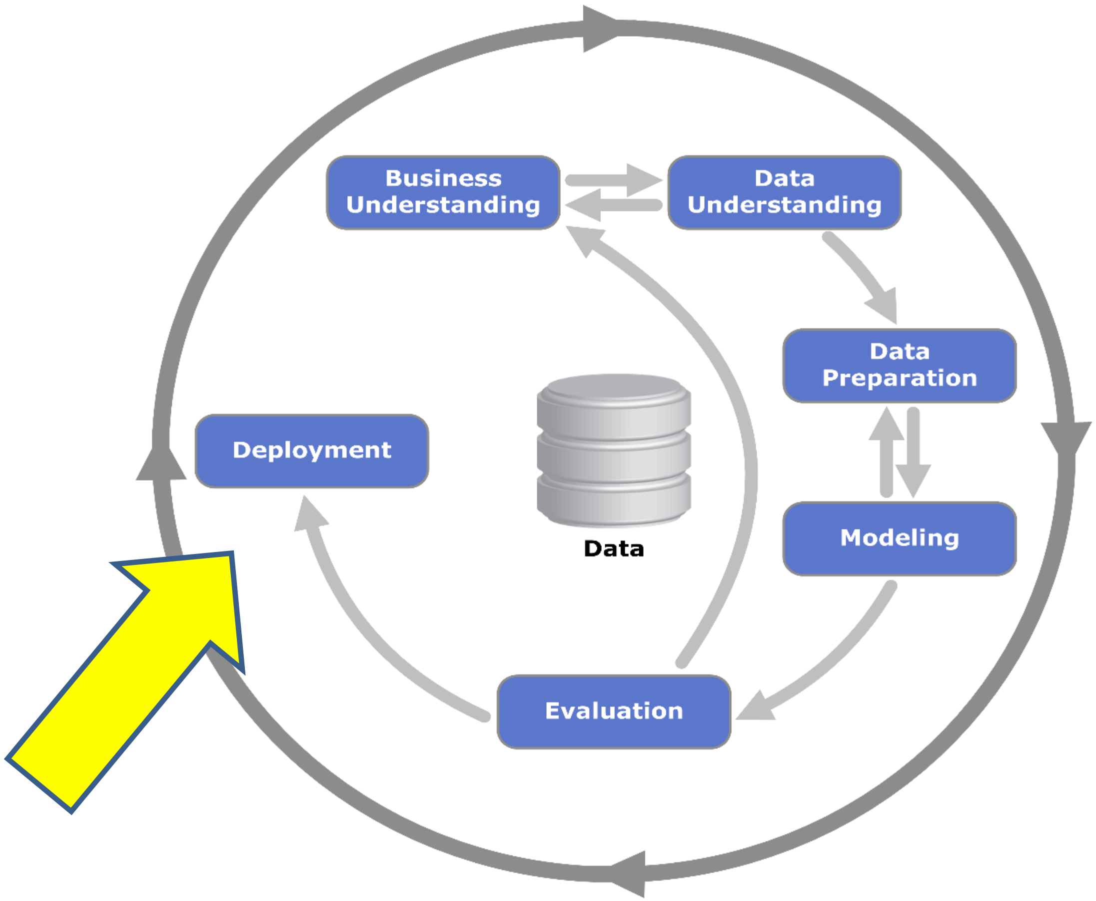


Neural Network



The process: CRISP-DM

Cross Industry Standard Process – Data Mining



Export the model to produce SAS code

Export Model

```
/*--------------------------------------------------
The options statement below should be placed
before the data step when submitting this code.
--------------------------------------------------*/
options VALIDMEMNAME=EXTEND VALIDVARNAME=ANY;
   /*------------------------------------------
   Generated SAS Scoring Code
     Date             : 23Sep2018:15:15:08
     Locale           : en_US
     Model Type       : Neural Network
     Interval variable: TurtleExactCombined
     Class variable   : Attack
     Class variable   : CrabLandingsDisc
     Class variable   : DissolvedO2discretize
     Class variable   : MoonPhase3daysextended
     Class variable   : pressurediscretize
     Class variable   : salinitydiscretize
     Class variable   : turbiditydiscretize
     Class variable   : windspeeddiscretize
     Class variable   : Direction
     Response variable: Attack
     ------------------------------------------*/   /*---------------------------------------------------------

   SAS Code Generated by Cloud Analytic Services for Artificial Neural Network

     Date                  : 23Sep2018:19:15:08 UTC
     Response variable      : Attack
     Number of nodes        : 51
     Number of input nodes  : 39

     Number of output nodes : 2
     Number of hidden nodes : 10
     Number of hidden layers: 1

     Type of neural nets    : MLP DIRECT

```

Export Cancel

Using the model

Intelligence from models can be operationalized to score new instances to see if shark presence is likely.

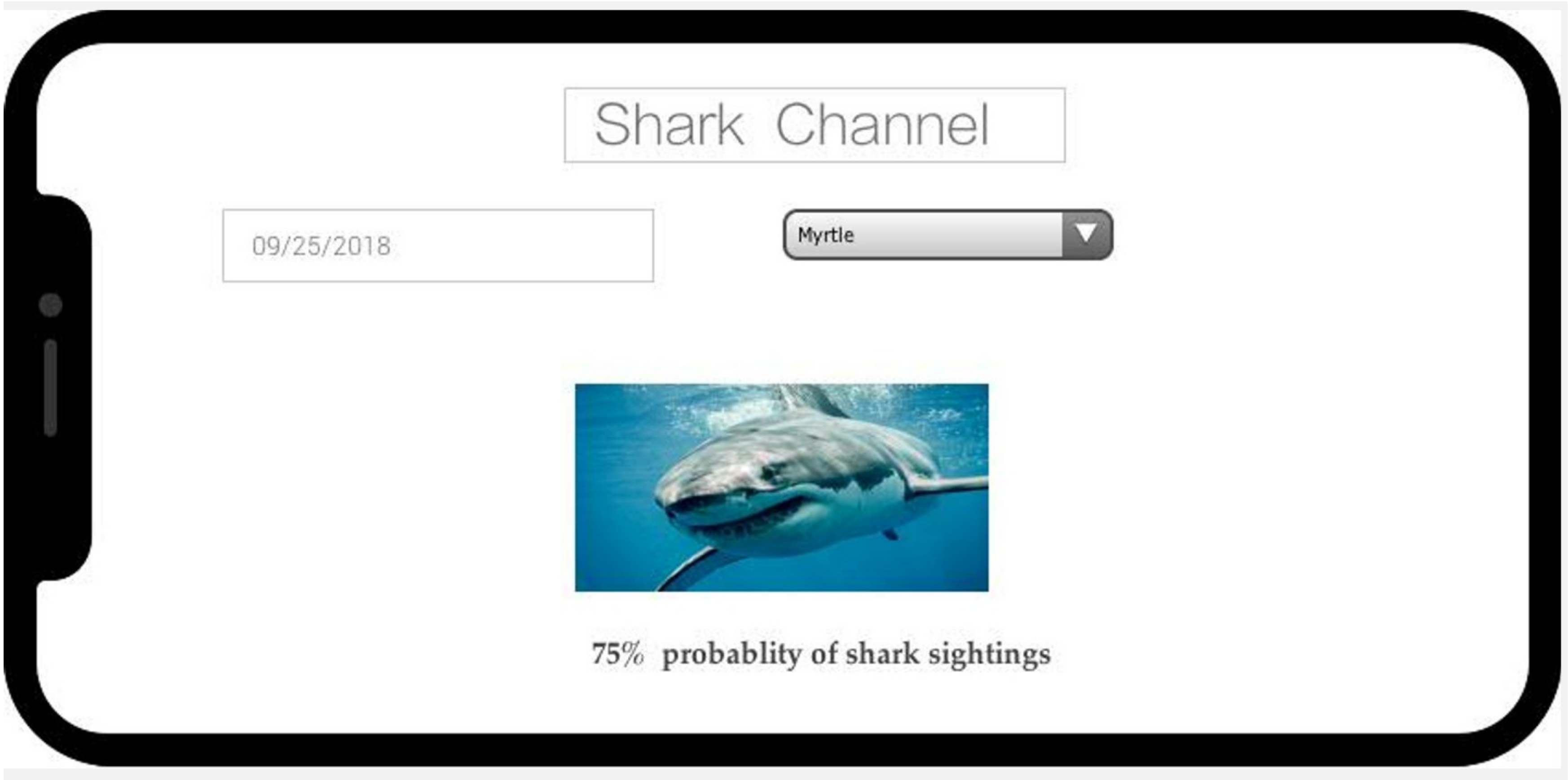
This model can then be delivered to the public in the form of an app to help improve beach safety. Additional data collected can further improve the model.

SAS Viya for Customized Mobile App

Shark Recommender App

The latest release of SAS Viya also gives you the ability to build and customize high-performance mobile apps for accessing content wherever you go.

Recommender System



UNC CHARLOTTE
The WILLIAM STATES LEE COLLEGE *of* ENGINEERING

Findings, Future Work

- Attributes of interest:
  - Weather: Wind Speed, Direction, Moon Phases
  - Environment: Crab Landings, Turtles
- Future Work:
  - Extend Data (drone, social media, other)
  - Collaboration with other researchers
  - STANDARDIZED HASH TAG
    - Promote Use for:
      - Shark sightings (see @sharkreports, @dorsalau on Twitter)
      - Turtle activity, Fish Schools
- Use research to promote STEM and Data Science to middle and high school students

UNC CHARLOTTE

#sharkchannel

“. . . investing in increasing our understanding of the behavior and ecological role of sharks as well as factors influencing the risk of shark bites, may be the most effective way to increase safety of people.

. . . .If people learn to avoid being near shark food during feeding times, we become far less likely to end up an as accidental appetizer.”

Dr. Francesco Ferretti, Stanford

References

Scientific American: Shark Bites are Up, But Attack Risk is Down? http://www.scientificamerican.com/article/shark-bites-are-up-but-attack-risk-is-down/

Dr. Chuck Bangley http://www.scientificamerican.com/article/shark-bites-are-up-but-attack-risk-is-down/)

Hashtag Standards for Emergencies: https://docs.unocha.org/sites/dms/Documents/TB%20012_Hashtag%20Standards.pdf

Lunar Cycle Effects: http://benthamopen.com/contents/pdf/TOFISHSJ/TOFISHSJ-6-71.pdf

What 3 Words: http://what3words.com/

Dr. Pamela Thompson blog: http://www.profthompson.net

Dorsal app: https://www.dorsalapp.com/

Sharktivity map and app: http://www.atlanticwhiteshark.org/sharktivity-map/

http://www.profthompson.net – Shark Research

https://peterjamesthomas.com/2017/10/11/hurricanes-and-data-visualisation-part-ii-map-reading/
Hurricane Prediction

https://pdfs.semanticscholar.org/e0c8/cd2ca2cf38afaf73b0265b6b5370a4066016.pdf

One study shows no effect: http://benthamopen.com/contents/pdf/TOFISHSJ/TOFISHSJ-6-71.pdf