The Maine Spaceport:

Rocketing Maine into the 21st Century

Data models for Maine's next Billion-dollar business

Roux Institute, Northeastern University 2021

Amanda Butler, Isabel Clarke, Alaina Cellini, Colin Kuhn, Daniel Griffin, Jeff Lavoie, Joseph Reynolds, Thomas Stahlhuth

September 02, 2021

The Maine Spaceport Initiative: Rocketing Maine into the 21st Century

Introduction

Maine’s natural resources and incomparable rocky coast have captivated the imagination of the world, but isolation and mill-industry decline have hampered the state's ability to retain, adapt, and draw new cutting-edge businesses. The Maine Spaceport Initiative seeks to change the tide by capitalizing on the fast-burgeoning “new space” industry-- a powerful combination of high-tech commercial space applications that include everything from engineering and launch services to ongoing software development and data analysis. Where space is concerned, Maine’s isolation is a geographic advantage that will create hundreds of high paying opportunities to grow the state's aerospace industry, STEM workforce, and educational opportunities.

The benefits of a spaceport in Maine extend beyond the companies who want access to the coveted polar orbits we can provide. Legacy industries of the Pine Tree State will certainly experience concordant opportunities. For example, it is easy to imagine how forestry services can better manage state and private lands with satellite-driven imagery, how endeavors in aquaculture and agriculture farming can gain efficiencies from remote monitoring, and how insights into tourism can be aided by an eye in the sky.

A spaceport provides Maine the opportunity to add a new tech sector, revolutionize its legacy industries, and even combat the growing threat of climate change, which helps us plan a more sustainable future for the next generation of Mainers.

With these possibilities in mind, Maine Spaceport Initiative has partnered with the Analytics program at the Roux Institute at Northeastern University to explore how satellite data could be leveraged and implemented by Maine's industries.

Statement of Purpose

Satellite data interconnects our daily lives in myriad ways we often overlook: cellular phone services, GPS, aviation, manufacturing, and shipping (to name a few) heavily rely on satellite data applications. Mainers are also connected by our legacy industries in practice and identity. Everyday we are surrounded by our forests, embraced by our waters and united by our collective identity as humble, hard workers and stewards of nature’s bounty.

A spaceport gives us the opportunity to live out these values in the 21st century and once again lead the country in the development of industry vital to innovation and infrastructure: “as goes Maine, so goes the Nation,” as the old adage says.

The purpose of this partnership is to explore ways that satellite data can benefit Mainers and the state’s legacy industries in order to help the Maine Spaceport Initiative communicate it’s value to stakeholders. The important insights gained through this work help to provide vision and perspective into the many facets of Maine's economy and culture while illuminating ways the Spaceport can guide its efforts to spearhead change and better the lives of Mainers-- and moreover, the lives of those beyond our borders. The Maine Spaceport helps Maine address 21st century problems by providing a direct route to emerging technology fields, updating our infrastructure, monitoring the impacts of climate change, and creating loads of new connections to a growing global marketplace; this project explores several ways the Spaceport can get liftoff.

Scope of the Project

The Maine Spaceport Initiative has partnered with the Roux Institute XN analytics course to identify use cases and analyze the impact satellite data can have for Mainers and the state's important industries. The Summer 2021 cohort working on the Spaceport XN project has identified three use cases for satellite data: (1) how to more efficiently allocate testing resources for aquaculture contamination, (2) how the state can improve monitoring and response to invasive species, using Eastern Ash Borers in Maine forests as a guide, and (3) how weather affects the capacity of rockets that the Spaceport can launch-- a meta parameter for the revenue potential and viability for an active spaceport on the coast of Maine (which is also modeled as part of the exercise).

The initial steps of the project were dedicated to identifying satellite data feeds and creating practical use cases for the Maine Spaceport Initiative. Public and private satellite data sources were explored to gain an understanding of the current landscape, and also to identify gaps.

Each group identified specific business problems, hypotheses, and business cases that could be solved with a combination of terrestrial and satellite derived data. The collection, preparation, analysis, and findings are presented for each use case as well as an initial attempt at creating machine learning algorithms and statistical models that could be used to make predictions. Lastly, each group addresses the subsequent iteration of analyses to be conducted by the next cohort.

Predicting Spaceport Revenue

Business Problems

Successful spaceports are not merely launchpads, but sophisticated consortiums of tech-heavy businesses that require extensive operational expertise that make them capital and labor intensive. In order for The Maine Spaceport to garner funding and achieve ongoing viability, it must be profitable. Concordantly, it is difficult to predict revenue outcomes based on little revenue data. However, while there is no record to draw from to create a model, our team was able to determine a weather-based launch cadence above, we know the potential of the market demand based on our research, and we have a price/Kilogram to put a rocket in the air, which triangulate enough data to begin to gain insight.

Hypothesis

Combine a predictive model for revenue simulation using launch provider and industry forecasts, to creating a picture of the Spaceport's financial feasibility over the critical growth period of its first few years.

Business Case

The Maine Spaceport needs to understand and communicate its revenue and profitability potential in order to effectively present its case to State officials, partner organizations, venture capital, and new tenants.

Background Research

Despite the extraordinary growth of new space in recent times, the sector is still considered an emerging technology, and as a result, small satellite tech is a specialized marketplace that does not yet have mass media coverage or an extensive peer-reviewed archive. There are, however, several reports such as Frost & Sullivan, Allied Research and SpaceWorks, that provide updated industry insight at regular intervals. These reports, along with interviews from launch providers, rocket engineers, NOAA officials, and Maine Government and civic leaders provided critical information.

Monte Carlo Simulation

This model seeks to be conservative in its estimates by framing the revenue potential of the spaceport based on a handful of key data provided by the proposed anchor tenant, BluShift Aerospace. The near-term (2 year) timeframe looks at the smaller, suborbital launch programs (mostly run by government, civic, and academic research entities), as well as a longer-term timeframe (5 years), that will open up projections to capturing 50% of the domestic micro-satellite marketplace within five years' time.

Revenue Projections

Historically spaceports have not been a profitable business. Some of the most established spaceports like the Mid Atlantic Regional Spaceport, or Cape Canaveral have not created profits in recent years even though they are considered to have an established launch cadence and have multiple anchor tenants. Although profitability may be a struggle, the ultimate sign of maintaining an active spaceport is having sufficient revenue to continue operations. The first few years are vital for the Maine Spaceport to establish its presence in the new space industry as well as create sufficient revenue from launch operations and its tenants.

Predictive revenue model & method

This model was built to be immediately informative and responsive to new inputs as The Maine Spaceport grows its footprint. Because there is no existing data to base assumptions on, parameters of the model were determined using data derived from the aforementioned market reports as well as from data provided by The Maine Spaceport's anchor tenant, BluShift Aerospace.

A triangle distribution model can be quite informative when working with only a few model parameters, and once coupled with a Monte Carlo experiment, the powerful union can iterate many thousands of outcomes, which creates reliable insights upon the aggregation of the experiments into a distribution. To that end we ran 10,000 trials of both the suborbital and orbital markets, providing visibility into the 2 year. and 5 year strategies at the Spaceport.

Findings

The outcomes of this study are quite favorable, demonstrating that The Maine Spaceport could quickly join high revenue-generating Maine industries like potatoes, cannabis, and Lobsters (fig. 1.08) to join the ranks as one of the biggest sectors in the state.

While near-term (2 yr) timeframe is hemmed in by a focus on the suborbital market as BluShift “proves its sauce” with smaller launch programs (mostly run by Government, civic, and academic research entities), the model suggests that revenue could range from $3M-6.5M per year during that time.

Once the company is fully open to the orbital launch markets forecasted by the research reports used to inform this study, the model points to the longer-term capacity of the Orbital market having the potential to generate between $250M-400M (5yrs) from this tenancy, as BluShift aims to capture 50% of the U.S. domestic micro-satellite marketplace within five years’ time.

Launch Cadence

Business Problems

The Maine Spaceport and industry professionals have identified Maines geographic advantage to launch satellites off the coast on a southward heading into polar and sun-synchronous orbits without the issue of potential overland flight that is experienced by other active spaceports along the east coast. But one thing they have not addressed is the launch capacity based on historical weather patterns, does the coast of Maine permit for a regular launch cadence? There are a range of variables that a launch provider must consider before conducting both suborbital and orbital flights.

Satellite data is a powerful tool that has changed forecasting modeling, with satellite constellations located and operated in low earth orbit by private and public entities like the National Oceanic Atmospheric Administration, SPIRE, and more. But how can the Maine Spaceport leverage this information to predict potential launch windows for their tenants and customers?

Hypothesis

By utilizing historical terrestrial weather data, we can effectively train a launch prediction model for satellite derived forecasting data.

Business Case

Currently, there is not a commercial service that offers a centralized platform that incorporates and monitors weather specifically for launching rockets. By creating a launch forecasting model, the Maine Spaceport will effectively be able to predict launch windows for anchor tenants, as well as attracting new launch providers and satellite manufacturers to the state.

Data Collection

Three sets of data were compiled in order to properly analyze historical weather patterns and predict future launch windows:

Data Set	Source	Location	Area of Observation	Observation Type	Time Period
1	Visual Crossing	Cutler, ME	Weather Stations	Historical Terrestrial	2010-2021
2	ForeFlight	Brunswick, ME	Executive Airport	Satellite Forecasting Data	8/11/21-8/21/21
3	ForeFlight	Bar Harbor, ME,	Hancock County Airport	Satellite Forecasting Data	8/13/21-8/18/21

Data set one will be used to address the first question in the business problem proving historical weather and its ability to consistently meet launch conditions over a ten year period. Data sets two and three will be utilized as testing sets for a support vector machine model that will predict three hour launch windows.

Research on Launch Weather Criteria

To properly prepare the data for analysis, research was conducted on historical and current weather needs to support launch. In the past, temperature was an important factor for launch criteria during the time of space shuttle launches. It was noted that a launch would not happen if the temperature was below 41 degrees for a 24 hours period. Wind speed and direction were also critical to launch both at the surface and upper levels of the atmosphere. Thunderstorms, cumulonimbus, lightning, were also restraining factors to launch (Dunbar 2003). For current providers like SpaceX and their falcon nine, similar weather criteria can be found. Wind speed can not exceed speeds of 30 knots 162 feet above the launch pad. Upper level conditions can not contain any level of wind shear. Cloud coverage, thunderstorms in a 10 nautical vicinity, cumulonimbus clouds, are all factors that call for a no launch scenario (NASA 2021). Although both of these launch vehicles are much larger than what is currently proposed in Maine, all attributes are vital for success.

Blushift Aerospace has been identified as a potential anchor tenant for the Maine Spaceport, a conversation was held with Sascha Deeri, CEO. During our conversation he identified the following characteristics: First, the wind speed must be blowing at a speed less than ten knots. Second, the sky conditions (cloud coverage) must be 50% or less. Third, wind direction can not have the potential for pushing the rocket back towards land. Fourth, the jet stream must be blowing less than 100 knots, and there can not be wind shear identified from altitudes 3k-50k. Fifth, these conditions must be met for consecutive hours in order to be approved by the FAA.

Data Preparation

Based on the research and our historical and forecasted data sets, the following parameters were created to predict launch conditions for Maine:

1. Wind speed < 10 knots
2. Temperature > 50 F
3. Sky Conditions = “Clear”
4. If wind is blowing from 60-210 and above 8 kts, do not launch
5. Three-hour windows that all parameters are met.

Due to historical weather data constraints, we were unable to include winds aloft information.

Exploratory Data Analysis Findings

Our analysis proved that from a time period of 2010-2020, historical weather did in fact support a significant amount of three hour launch windows. On average there were 2,169 opportunities to launch over this ten year period (Figure 1.03). Hypothetically, this could support the revenue projections found in the monte carlo simulation analysis.

Due to the constraint of temperature, launch windows are rare in the months of December, March, and April. But, there is a significant portion of opportunities in June, July, August, and September (Figure 1.02). The Maine Spaceport will have to account for this information in their business model as well as potential anchor tenants and their specific launch condition needs.

Our analysis did not place parameters on the time of day for the three hour windows, so hypothetically launch windows could exist from 12am-3am, 8pm-11pm, and so on. We were able to identify that historically the afternoon 1pm-4pm offered the least opportunities for launch (Figure 1.01).

Of the variables utilized to create our launch decision variable, wind speed and air temperature will be the main inhibitors to launch off Maine's coast with micro launch vehicles (Figure 1.04). If the Maine Spaceport were to decrease the restriction of air temperature, and find launch providers capable of handling increased wind speeds, a significant portion of launch windows would become available, increasing revenue opportunities.

One of the last factors that needs to be accounted for is the wind direction of the Cutler region. Wind blowing in a direction of 60 to 210 degrees with speeds above 8 knots have the potential to push rockets back towards land. Our analysis has shown that wind does not tend to be a significant factor in directions 91-210 degrees. The majority of the wind can be found in the 271-360 direction which is ideal for launch (Figure 1.07).

Launch Cadence Support Vector Machine Model

Why Use SVM?

Understanding and predicting conditions is critical to avoid catastrophes that could cause damage to the spaceport, cargo, or ultimately human life. Either all conditions are met for a good launch, or they are not. Launching a rocket can be categorized as a binary outcome: launch, no launch. The support vector machine is a supervised machine learning algorithm known for its strength in binary/classification based problems. It is also known for its ability to limit overfitting that is observed in other machine learning algorithms.

Data Sets and Variables

The following information categorizes the models variables, training data, testing data, and the total number of observations for each set of data.

All model variables are known for their binary outcomes of either meeting or not meeting the launch criteria. 0 = no launch and 1 = launch.

Model Variables:

1. Wind_Go
2. Temp_Go
3. Conditions_Go
4. Wind_Direction_Go
5. One_Hour_Window
6. Launch

Data Set	Train/Test	Variables	Observations
Historical Cutler	Train	6	99,678
Forecasted Brunswick	Test1	6	231
Corecasted Bar Harbor	Test2	6	111

Our model is attempting to accurately predict forecasted three hour launch windows based on historical weather patterns. Our prediction value will be known as “Launch”.

Model Results

All three models tested above a 90+% accuracy level as seen in the table below.

Model Data	Accuracy
Train	93.24%

Test1	96.54%
Test2	94.59%

Confusion Matrix for Model Performance

But upon further analysis of the confusion matrix for model performance, we were able to identify multiple false negative errors(highlighted). This is considered a type two error where a go launch window is actually predicted as a no go window.

Train	Predicted Value		
Actual Value		True (0)	False (1)
	No Launch(0)	68,670	0
	Launch (1)	6,733	24,275

Test1	Predicted Value		
Actual Value		True (0)	False (1)
	No Launch(0)	28	0
	Launch (1)	8	195

Test2	Predicted Value		
Actual Value		True (0)	False (1)
	No Launch(0)	17	0
	Launch (1)	6	88

The launch cadence analysis was able to prove historical viability for supporting suborbital and orbital launches off the coast of Maine. The estimates used for creating the three hour launch windows were conservative estimates that were tailored to micro launch vehicles like Blushift Aerospace. The Maine Spaceport is able to change the variables based on the needs of launch providers that could potentially create more historical launch opportunities. Our model has proven the power of utilizing historical terrestrial data to effectively train a model that can use satellite predictive data to predict launches correctly with 90%+ accuracy. Steps will need to be taken to correct the type two errors to avoid potential misidentification of forecasted launch windows.

Next Steps for Predicting Launch Cadence and Revenue

Further research and data collection will be needed to create a finished launch prediction model for the Maine Spaceport Initiative. Two key variables that were not included in the model were the winds aloft and the cloud ceilings. We have identified the following sources for each variable.

Sources for Winds Aloft Data
1. ForeFlight: Surface Level - 51k MSL
2. NOAA Aviation Weather Center: Surface Level - 53k MSL
3. Spire: Surface Level-100km AGL
4.European Center for Medium-Range Weather Forecasts

Sources for Cloud Ceilings
1. NavLost (historical METAR Data from AWOS/ASOS stations in Maine
2. NOAA Aviation Weather Center
3. European Center for Medium-Range Weather Forecasts

Current terrestrial observation systems limit the height and area that can monitor meteorological conditions that are critical for launch. Low earth orbit constellations like Spire have the capability of increasing weather observations from the surface to high levels of the atmosphere while covering the entire geographic launch area and identifying all weather variables from a centralized platform.

Maine Forests & Invasive Species

Business Problem

Invasive species, such as emerald ash borers, pose a serious threat to ash trees and the forestry industry. Today, identification of these species is done by land sampling, surveying, and estimation. This field work is completed by state forest services and is manual and resource intensive. Invasive species and pests can have a massive economic impact on the forestry industry and the environment. Improved data collection processes such as remote sense can be used to better track invasive species and predict their spread.

The invasive species our research is focused on is the Emerald Ash Borer (EAB). The EAB was first identified in Michigan in 2002. It was later identified in Aroostook County in 2018 and in York County in 2019. Now, the EAB has been discovered in 15 percent of the counties in the 37 states that have native ash trees. EAB research in Michigan is focused on forest degradation and a mortality lag of 4 to 7 years was identified. Due to this lag, Maine’s forests will not see an increase in tree mortality until 2022. The problem we’re aiming to solve is not to understand the areas that have degraded but rather to predict those areas most at risk for EAB infestation.

Solution

Using forest inventory and Emerald Ash Borer (EAB) infestation data, we have built a model that can identify geographic areas at most risk of infestation. This model has the potential to be improved through the use of satellite imagery or remote sensing data. This type of data would work alongside field surveying which is necessary to understand the invasive species and characteristics to look for when remote sensing is layered in. The EAB case data we received from the State of Maine's Department of Agriculture, Conservation and Forestry represented 53 trap locations across the state. The coordinates of these traps are from both purple and girdled tree trap types and the actual locations have been buffered to protect landowner privacy. The forest inventory data is limited to plots that have been surveyed and some of these plots map to multiple positive EAB cases. Given this limitation, satellite data would be a better source for the forest inventory input as the model will not be limited to plots that have been sampled during the annual FIA surveying process, which is only 10 - 20% each year.

Business Case

By using forest inventory data, we can identify areas at risk for EAB infestations to support pest and forest management efforts and policy. In fact, on August 30th, 2021, the Maine Forest Service announced an emergency order restricting the movement of ash tree materials. The policy is intended to slow the speed of EAB infestation and buy time to figure out ways to limit the impact of EABs in Maine's forests. Today, 90% of the state's ash trees are outside the area of regulation. The ability to predict areas at risk is crucial to effectively regulate the spread of these pests.

Forest Inventory data is collected via sampling and estimation. Thus the model is limited to plot data that has been collected in this same way. The accuracy of the at-risk plot classification model would increase through the use of satellite data as data coverage would not be reliant on recently-surveyed plots.

EDA

Data Collection & Prep

The Forest Inventory and Analysis (FIA) program of the USDA Forest Service has amassed data about the forest's of the United States since 1930. Their mission is to:

> ***"Make and keep current a comprehensive inventory and analysis of the present and prospective conditions of and requirements for the renewable resources of the forest and rangelands of the US."***

The FIA datasets compile statistics on a national level but also on a state level which then can be subset to county level and or per plot level; and if we needed to get very granular we could see all of the data collected on a per tree basis from every survey plot. In reviewing this data we found many ways to access it and once downloaded there are 70 different tables (some at the state level and some used as reference tables) that can be accessed. The data covers surveys completed across the State of Maine (Fig. 2.01). There are 18,044 unique plots that FIA have surveyed since 1995. Out of these we were able to identify 1,309 or 7.25% of the survey plots have Ash trees on them. The quantity and time-based availability of the data is particularly important for the EAB infestation given the four to seven year delay between identification of the EAB in a forest and the impact to tree mortality and volume.

Data Preparation

In the FIA dataset we focused primarily on two tables; Plot and Tree. When subsetting to just Maine we were able to find the exact location (via latitude and longitude) of the survey plots and review all of the rollup up statistics about these plots. For example; what county is it in, does the plot have water, what is the size of the plot, categorization of the plots by ecological characteristics, ect. When joined to the Tree dataset we were able to look at the unique characteristics of each plot and see the make up of trees by species, size of the tree (height and diameter), condition of the tree, ect.

In addition to exploring the FIA datasets we were fortunate to acquire survey data from the State of Maine's Department of Agriculture, Conservation and Forestry that is tracking positive identification of EAB infestations in the State. As noted above some of the data's longitude and latitude coordinates were "blurred" to keep the privacy of land owners. To map/match the EAB positive sites to the FIA data we needed to identify if the EAB positive sites were on any of the FIA survey sites. In our matching we did not find many positive matches and refocused our efforts on identifying the nearest neighbor based upon latitude and longitude of the EAB positive data to the FIA plot data set. From this matching we can start to identify the characteristics of the plots that are ideal for EAB infestations.

EDA Findings

Overall, tree abundance in Maine has remained relatively stable over the last 20 years (Fig 2.02). This indicates that over harvesting and over development has not been an issue here in Maine. What was interesting was the abundance of standing trees by county, where in fig 2.03 we can clearly identify that Pisctaquis county has the most standing trees per acre. This might be the case where conservatorship land was gifted to the federal government in 2016 to expand Baxter State Park. While not part of this EDA land ownership may play a role in abundance of standing trees per acre as well as population density per county that directly impacts standing trees per acre.

We learned there is variance in forest density and abundance by county (Fig. 2.03) and that ash trees do not populate the entire state. Out of the FIA me_tree dataset there are 90 different species of trees in Maine where 25.4% of the total survey tree are Balsam fir followed by Red Maple, Red Spruce, Northern White Cedar, and Paper Birch at 11.8%, 10.7%, 7.6%, 5.9% respectively. Out of the total survey population Ash trees account for only 1.95%. The majority of ash tree coverage in Maine comes from the white ash species and are in the size range of 10 to 14 inches in diameter (Fig 2.04). As stated above out of the 18,044 FIA survey plots only 1,309 have ash trees on them which is not surprising given the low percentage of total population in all survey plots. In fig 2.05 we introduce the estimated basal area by county for ash trees which shows a cross functional heat map of where ash trees grow and what size are they. This map highlights that the larger ash trees populate in Sagadahoc, Lincoln, Cumberland, Androscoggin & Waldo counties.

Within the FIA plot dataset the survey plots are categorized by ecological subregions. The categories link similar surficial geology, lithology, geomorphic process, soil groups, subregional climate, and potential natural communities. As described by FIA the boundaries of each subregion abruptly change and correspond to discrete changes in geomorphology. Within Maine there are 16 ecological subregions. Out of the total surveyed plots in Maine the following are the top 5 proportions that account for 50.1% of the total plots. Figure 2.07

Northeastern Mixed Forest	Adirondack-New England Mixed Forest	Northeastern Mixed Forest	Adirondack-New England Mixed Forest	Northeastern Mixed Forest
Central Maine Embayment subsection (211Da)	St. John Upland subsection (M211Ab)	Maine - New Brunswick lowlands (211Bb)	Maine Central Mountains subsection (M211Ac)	Aroostook Hills Lowlands (211A)
13.9%	12.3%	10.4%	7.9%	5.6%

Within the 16 ecological subregions the top five that have more Ash trees in the survey plots were the following;

Adirondack-New England Mixed Forest	Northeastern Mixed Forest	Northeastern Mixed Forest	Adirondack-New England Mixed Forest	Northeastern Mixed Forest
St. John Upland subsection (M211Ab)	Central Maine Embayment subsection (211Da)	Aroostook Hills subsection 211Ab)	Maine Central Mountains subsection (M211Ac)	Aroostook Hills Lowlands (211Aa)
13.7%	12.9%	11.4%	8.3%	8.7%

In reviewing the EAB positive finds data we noticed these data points are congregated in southern Maine, York and Cumberland county, and a few positive finds in Northern Aroostook county at the border of Canada. This is not surprising as the infestation in other states has started well before the 1st find within Maine in 2018. Additionally positive infestations in Canada make Maine one of the last Northern states to be impacted by the EAB infestations. The lack of data from EAB positive finds was a challenging aspect of utilizing the FIA data sets to identify commonality between survey plots. There seemed to be randomness in the positive infection sites with no linear or close proximity. Along with movement of firewood between state lines, by campers and vacationers, may pose a problem in identifying the next positive finds as well as having trained eyes spot infected trees.

Classification Model

Why a classification model?

To model the threat of the EAB to certain areas of Maine's forests we used a decision tree classification model. The data for the EAB was limited to 47 records dating back to only 2018, so a time series analysis could not be used at this time. Given the limitations to our data, a classification model is the best option. The outcome our model is looking for is whether or not the existence of an EAB for a given plot is True or False.

Data Used

From the FIA data for Maine, the variables plots, conditions, and ash trees were combined. Only the data from 2018 and 2019 were used, since a lot of the plots are resampled every few years. These years also match when the EAB first appeared in Maine, which allows us to match the discovered EABs with specific plots. The closest plots to each EAB were calculated, and the EAB condition was mapped to each of those plots. The number of plots with EABs was fewer than the number of EABs found, since many of the EABs were found close together and their distance between plots was greater. This reduced the number of locations with positive finds to 25.

Sampling Techniques

The data was split 80:20 for a training and test set to create a decision tree classification model (Fig 2.08). There were four splits with the variables ecological subsection code, net cubic-foot volume, measurement month, and net annual merchantable cubic-foot growth of a growing-stock tree on timberland. The model achieved an accuracy of 98% which usually would be considered a successful model, but the prevalence of the classier was 99.2%. Since the accuracy is lower than the prevalence, it is worse at classifying a plot than if you said all plots were negative for potential EAB sites.

A variety of balancing methods were used including oversampling, undersampling, a combination of both, and the ROSE method. All of these methods, except the combination method resulted in a lower accuracy measure, and the best one was still lower than the prevalence. The balanced accuracy scores of all the models fell below 80%. The combination of sampling methods model had 4 four splits using ecological subsection code, net annual merchantable cubic-foot growth of a growing-stock tree on timberland, total height, and compacted crown ratio (Fig 2.09).

Findings

We found these modeling techniques to be unsuccessful, and attribute it to the lack of plot precision and EAB data. Models from other states could potentially be used to identify features, if their forests share commonalities. Additionally, using satellite imagery would allow a closer inspection of the conditions in the exact geographic coordinates the EABs were found, and give 47 plots to pick from.

Conclusions & Next Steps

As discussed, there were a handful of limitations our model had to face. First is the use of sample data to represent the forest inventory in the state of Maine. A satellite imagery feed would be a better fit as the EAB case data would be able to match up exactly with each geographic area and not simply the closest plot. There is also a limitation to the EAB data itself. We had access to 47 positive cases from the state of Maine, all of which were located at the northern and southern borders of the state. Given the size of this data set, we were facing imbalanced classes for our target variable and tested a number of sampling techniques to improve the accuracy of our model. We were also unable to map all 47 positive cases to separate forestry plots given the limitation of the Forest Inventory data. Joining satellite imagery data to the EAB dataset would allow all 47 cases to map to a geographic area which would move the sample closer to being more balanced. We also had access to areas with traps but no positive cases, better inventory data would allow us to look just at those areas with traps and further identify their geographic characteristics.

Gulf of Maine Shellfish

Introduction

Maine began monitoring shellfish harvesting areas and publishing coastal water sampling data in 2013. In the trailing eight years, 97.54% of 55,675 samples were taken at random across 1,217 distinct sites. Within the samples, fecal coliform per 100ml of seawater scores range from 1.5 to 1,700 with 70.48% of scores equal to two or less. Less than one percent of all samples taken are “investigatory,” but this method has a fecal score distribution that is significantly higher than samples taken at random (Fig. 3.01). In theory, increased targeted sampling will result in more meaningful physical inspections and help prevent inadvertent harvesting of high-risk areas.

Business Case

To determine if a machine learning model can accurately forecast contaminated areas to increase impactful investigatory testing thereby reducing testing costs, health risk and lost business revenue.

Data Preparation

The initial shellfish data obtained from the State of Maine GIS catalog was an aggregated testing sample, by site, with a minimum date representative of the date the first sample of the aggregate was taken. Oceanic metrics were joined to this dataset by pulling National Oceanic and Atmospheric Administration's (NOAA) buoy data from six locations and joining daily metrics to the nearest testing site by minimum testing date. Despite extensive coding, exploration within this newly established dataset exposed little correlation between oceanic metrics and fecal contamination scores.

Logically, the lack of correlation was inconsistent with our initial premise that contamination thrives in warm water. We speculated that the combination of aggregated sampling data using a single date and overall distance between sample and weather buoy was hiding potential correlations and opportunities for an accurate model. In an effort to improve underlying data we contacted Bryant Lewis, the Growing Area West Program Supervisor for the Maine Department of Marine Resources. Bryant was able to provide an expanded set of granular testing data. This new dataset of individual sample metrics included contamination score, temperature, salinity, tide level and other categorical data such as wind direction and if the sample was taken by boat or land. The most significant challenge of working with this new data was related to considerable skew within the fecal contamination scores (Fig. 3.02).

Secondary Dataset Cleaning

The state provided data was broken into two separate "grow areas" called "WA_WZ" and "EA_EU." A union was utilized to create a single point of data. The data did not include a latitude and longitude, but it did provide a unique station identifier. Using data available in the online GIS data catalog, a join was utilized to connect the latitude and longitude to the testing data by this identifier. Much of the naming conventions were vague and were changed using metadata provided by the state. For example, the sample method of "L" was changed to "Land" and "B" to "Boat." Temperature values taken on 05/27/20 were taken in error and were corrected. It appears that the tester logged the temperature in Fahrenheit, but the data column was Celsius. We also created new columns such as a Celsius to Fahrenheit conversion, a top quartile fecal contamination identifier and a salinity > 24.5 identifier. Minor structural changes were also made for time series analysis in R. Further preparation is to consider statistical methods such as discretization and winsorizing to normalize sections of the data for further analysis.

Findings

Our initial examination of the data was to determine if temperature is directly related to fecal contamination score. To find this potential correlation, both mean temperature and scoring data was aggregated by month, converted to a time series and plotted (Fig. 3.03). This plot reveals a seasonal relationship between temperature and score but a scatterplot (Fig. 3.04) removed any notion that significant correlation exists.

A decision tree algorithm was utilized to determine if any additional variables may have predictive qualities. This identified a salinity level of 24.5 or less as a possible intersection to consider for further analysis. Additionally, the adversity rating of “P” or “Precipitation noted within the last two days” is also of potential interest. Starting with salinity, it was determined that when salinity is above 24.5, the median fecal score is 1.9 otherwise, the score is 4.0. Specific to precipitation, the median variance is slight at 2.0 versus 1.9 but the mean of 30.2 is nearly three times more than the 10.9 that exists when precipitation is not present.

A spatial analysis was considered to determine if high scoring sites could be identified geographically over time. Using Tableau, a plot was built to identify areas of low salinity and top quartile fecal contamination scores when precipitation occurred in the previous two days. Theoretically, this would help to determine what areas are most susceptible to contamination and when. For pattern analysis, data was segmented into four maps showing fourth quarter results over the trailing four years (Fig. 3.05). The lack of overlapping top quartile contamination sites suggests that issues are either mitigated when discovered, or the amount and location of rainfall is markedly random.

The above analysis established a need to contact the state again to determine what might be causing the lack of spatial correlation and reduced fecal contamination scores year over year. For the former, Bryant confirmed that when residential contamination sources are identified, the state works with town plumbing inspectors to compel property owners to address contamination sources. This type of impact can be seen throughout the whole state but Vinalhaven island best shows the effect to a small area (Fig. 3.06) and (Fig. 3.07). For the latter, Bryant advised that sampling methods have also been altered to incorporate more boat testing in lieu of land tests each year (Fig. 3.08). Boat samples on average trend below land samples for fecal contamination scores (Fig. 3.09). These factors matter because building a model is dependent on past history repeating itself, in the case of shellfish data, the present is being altered, which will inhibit prediction of the future.

Modeling Analysis

Regression models were ruled out early due to the overall lack of numerical correlation within the variables. Overall, a linear trend did exist between temperature, salinity parts per thousand and fecal contamination scores (Fig. 3.04) and (Fig. 3.10) but, the residuals were too large for accurate prediction.

Cluster analysis identified potential for a decision tree model which was run using an 80% training, 20% testing split. The results of the model (Fig. 3.11) identified salinity, temperature, growing area and “Adversity” as splits. Adversity categories are defined as followed: P = “Precipitation: Rain or mixed precipitation anytime within past 2 days (i.e. thunderstorms, rainfall more than a drizzle), and N = “Nonpoint: Flowing streams, stormwater pipes, or overland runoff” and W = “Wildlife: Waterfowl (10 or more), domestic or wild animals (i.e. at the station or in close enough proximity to have a possible impact)” (Lewis, 2021). The categories of “9” and “R” are undefined.

The model confusion matrix (Fig. 3.12) exhibits an overall model accuracy of 78.2%. However, this result was misleading due a large concentration of low scores being correctly identified in the model. In this case, when the score was actually low, the model predicts it correctly at a rate of 98.05% (8,396 of 8,563). Conversely, when the score was actually high, the model only predicted it correctly only 10.28% (257 of 2,499) of the time. Where the model does offer some value to the state is relative to this high prediction in general. When the model predicts a location to have a high score, it will be correct 60.61% of the time. Recall that the predominant testing method for the state is random 98% of the time. Of that, they should expect to see around 22% of those inspections result in a high score. If the state focused on the small subset identified as "high," this high score rate would increase nearly threefold to 60.61%. In the trailing three years, the state has conducted only one targeted sample out of 19,765 samples. This method would increase targeted sampling to approximately 256 per year.

Conclusions & Next Steps

Due to mitigation efforts and changes in sampling techniques, historic data is unable to make an accurate predictive mode for shellfish contamination sites. This highlights the need for real time satellite data to accurately predict areas at risk for contamination. Our analysis shows that historical data can be used to model areas at risk for contamination and improve testing efficiency. However, the accuracy of the model is hindered by mitigation, changing sampling methods, and lack of focused data. By utilizing satellite imagery and remote sensing technology, a more accurate model can be built using the variables in our model such as temperature, salinity and precipitation, as well as additional variables provided by satellites such as chlorophyll concentration and land use near the coast. Not only will satellite data improve the shellfish industry by accurately and efficiently tracking contamination, but could also be used to locate optimal new harvesting sites and lessen the impact of pollutants with real time detecting.

Final Conclusions

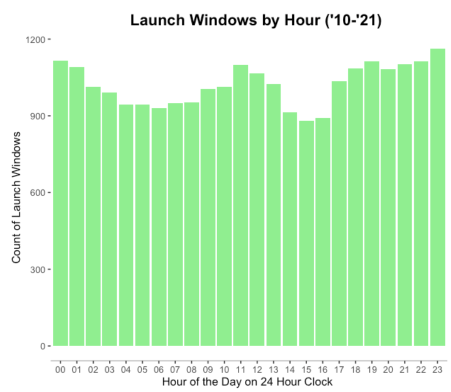
The Summer 2021 partnership between The Maine Spaceport and Roux Institute produced strategic insights that highlighted some of the benefits satellite data will have for Maine's legacy industries, simultaneously demonstrating the Spaceport's viability and potential economic impact. This research will help the Maine Spaceport Initiative communicate it's value to stakeholders and spearhead actions to further the availability of advanced satellite data in Maine.

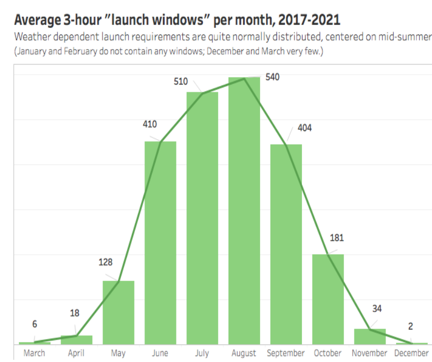
Aligning academia and industry with the growing technological trends in the new space economy, The Maine Spaceport and the Roux Institute are uniquely poised to support the state's heritage industries while growing the application of limitless possibilities that can only be recognized by communities with a diverse technological ecosystem. Emerging opportunities to combat climate change, monitor smart cities/utilities, public health, safety, security, remote assets, search and rescue, and natural resources are the tip of the iceberg.

At this checkpoint, the team is pleased to have uncovered use cases that point the way to increased efficiencies for risk management within the Forestry and Aquaculture sectors and identified specialized skills and satellite data that allow for even more refinement of the models that were developed. While our XN team regrets there were not another eight weeks to work on this project, we know that the next group will follow behind us and pick up the torch in order to continue the exploration of new space with The Maine Spaceport.

Appendix

Launch Windows in Weather and Revenue Forecasting

<table>
<tr><td>Figure 1.01 - Total launch windows by hour 2010 - 2021</td><td>Figure 1.02 - Total launch windows by Month 2010 - 2021</td></tr>
<tr><td>


</td><td>


</td></tr>
<tr><td>Figure 1.03 - Launch windows by Year</td><td>Figure 1.04 - Total instances of wind direction by compass degree 2010 - 2021</td></tr>
</table>

3 Hour Launch Windows ('10-'20)

Year	Launch Windows
2010	2337
2011	2230
2012	2273
2013	2119
2014	1951
2015	2131
2016	1978
2017	2276
2018	2029
2019	2318
2020	2220
	Annual Average = 2169

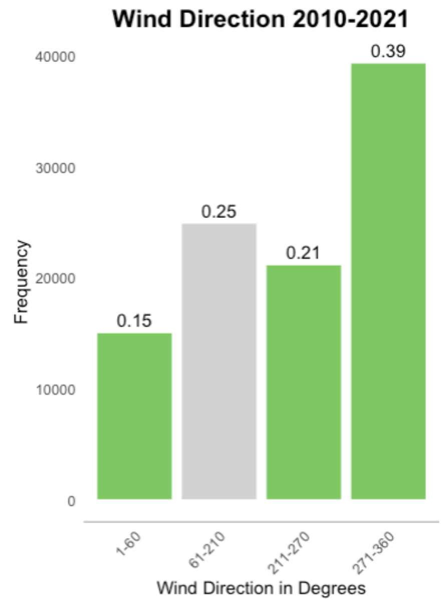
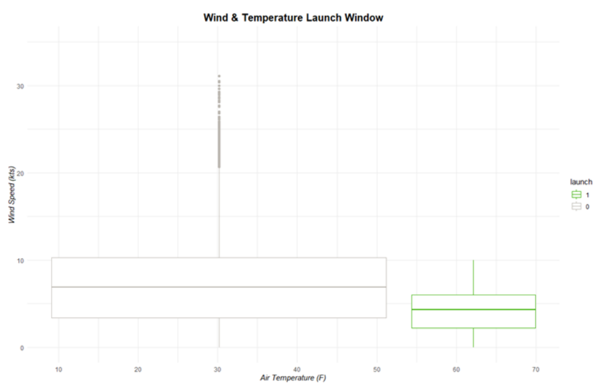


Figure 1.05 - Windspeed / Air Temperature Launch Window boxplot

Figure 1.06 - Linear correlation between wind speed and air temperature



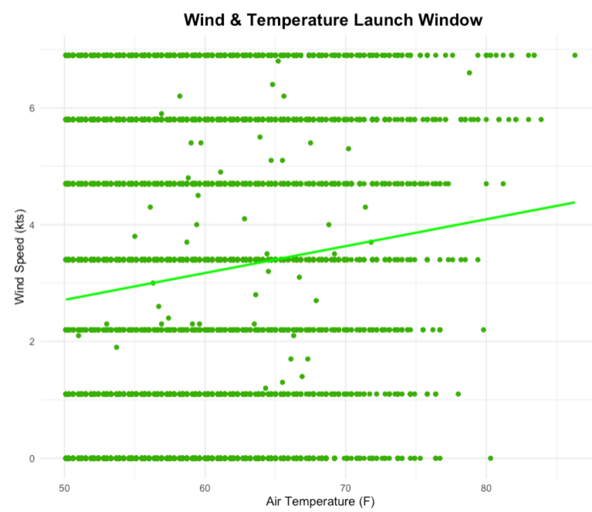
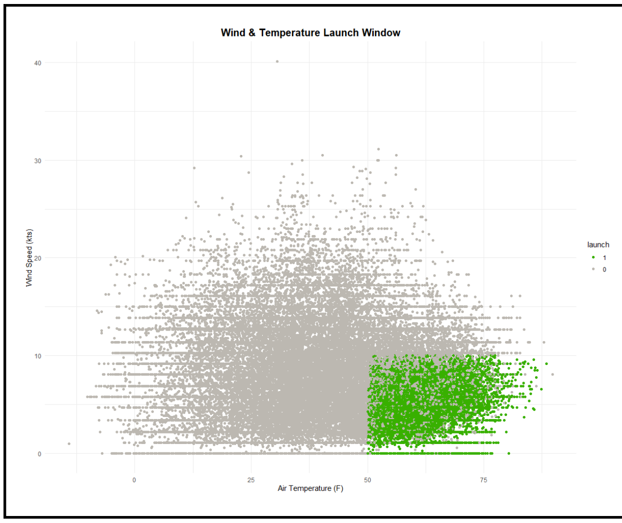


Figure 1.07 - Scatterplot of launch windows that meet the wind speed and air temperature launch parameters

Figure 1.08 - How the Spaceport revenue predictions stack up to Maine's biggest industries



Is space Maine's new big industry?	
Industry	Revenue
Lobster	485 M
Potatoes	167 M
Cannabis	112 M
Spaceport	***55-60M***

Impact of invasive species to Maine forests

Fig. 2.01

Coverage areas of ME FIA data

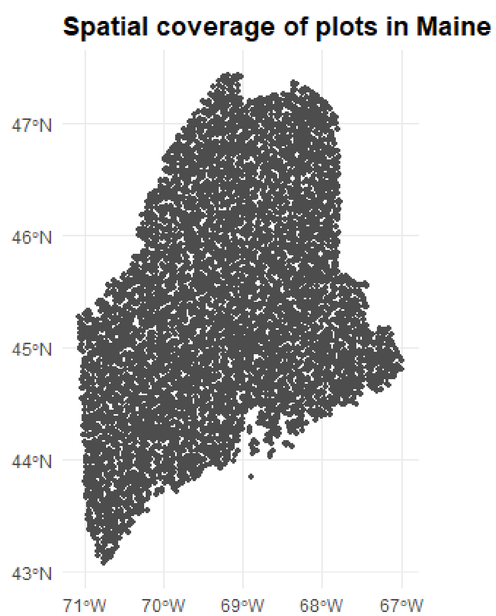


Fig. 2.02

Estimated abundance of standing trees through 2020

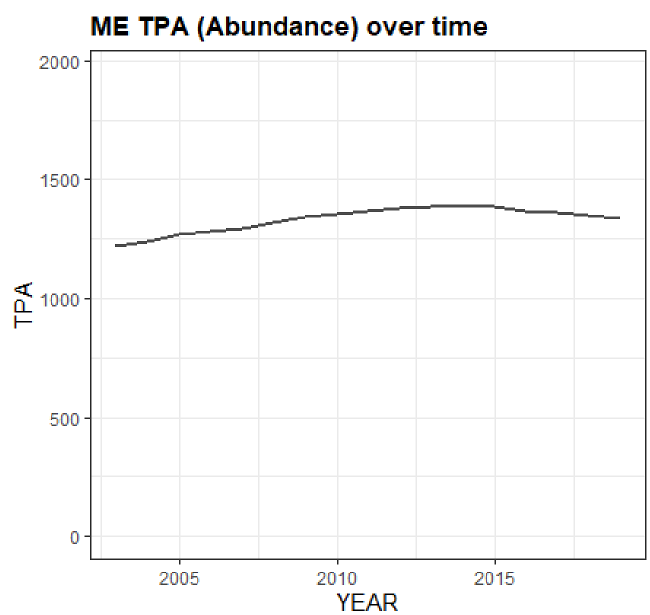
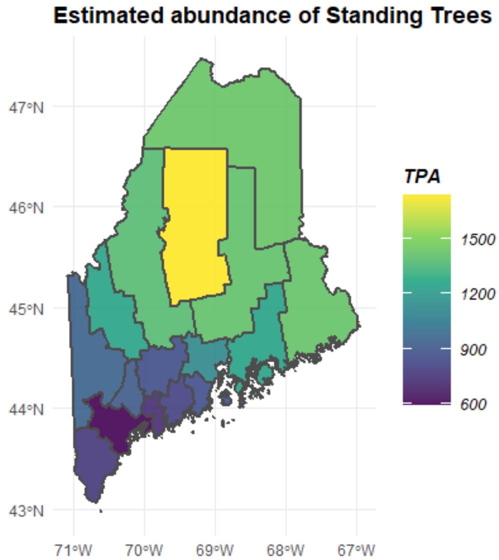


Fig. 2.03

Fig. 2.04

Standing trees by acres and county



Size distribution of Maine Ash trees

Fig 2.05

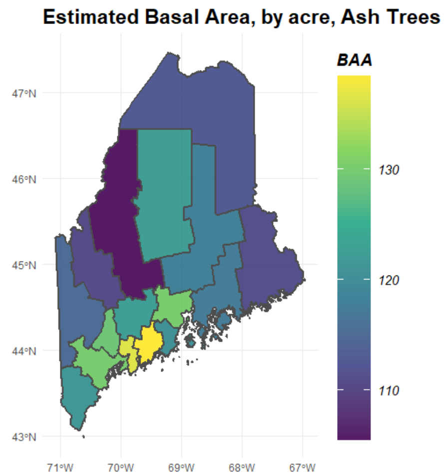


Fig 2.06

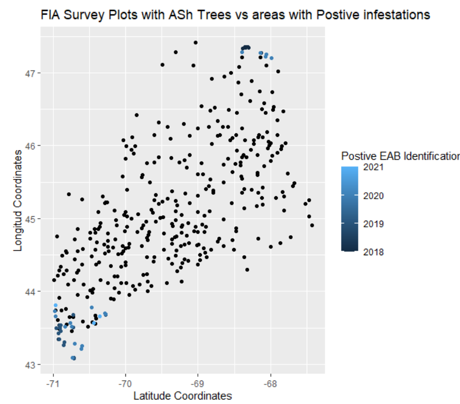


Fig 2.07

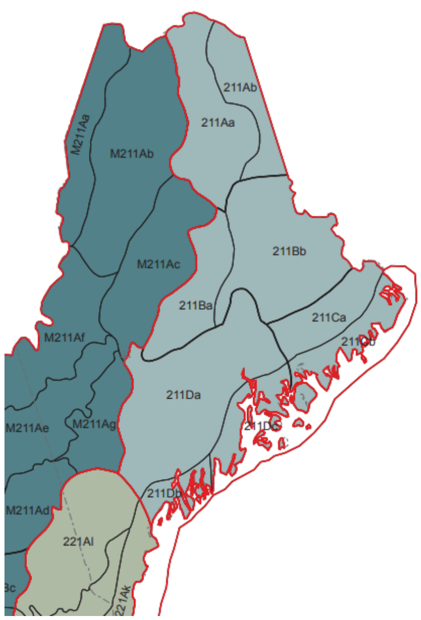


Fig 2.08

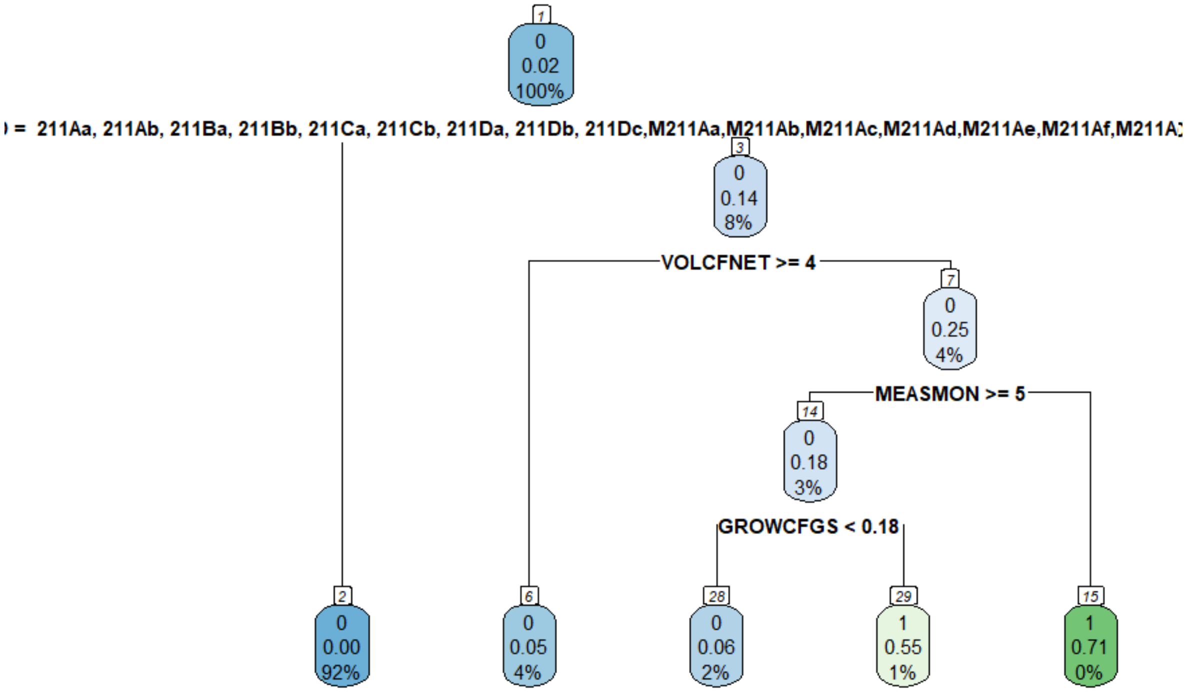
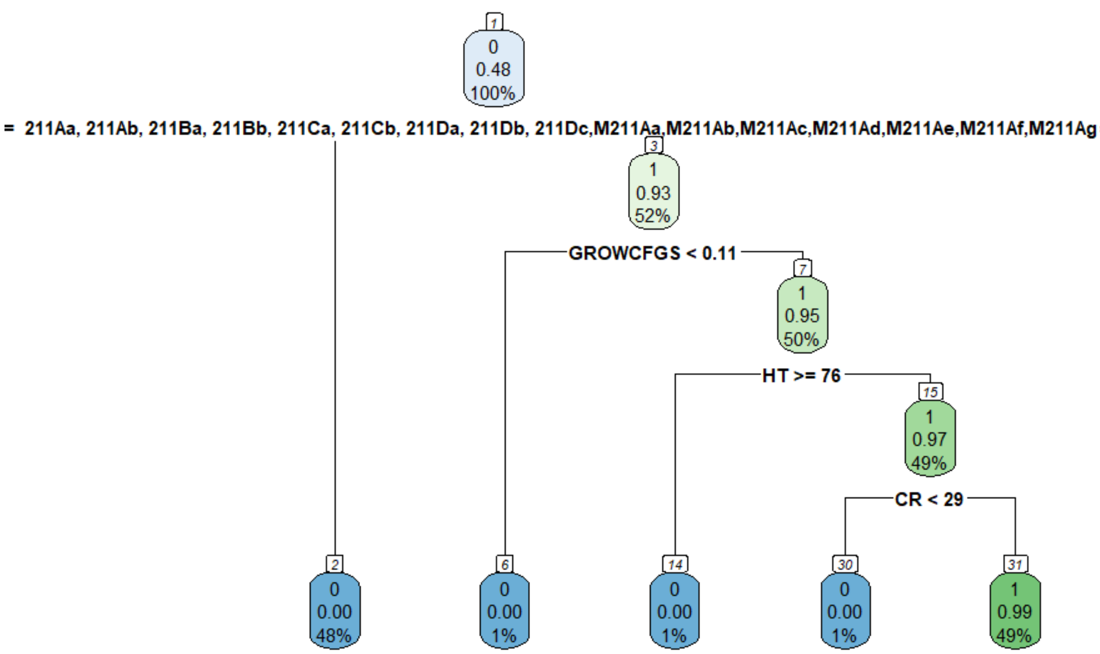


Fig 2.09



Shellfish EDA Findings and Modeling Analysis

Fig. 3.01. Distribution of fecal scores when samples taken at random against investigatory. x Axis altered to show distribution as significant right skew inhibits ability to visualize distribution. x Axis is relative to density. The axis numeric are not essential as this is just to show distribution shape

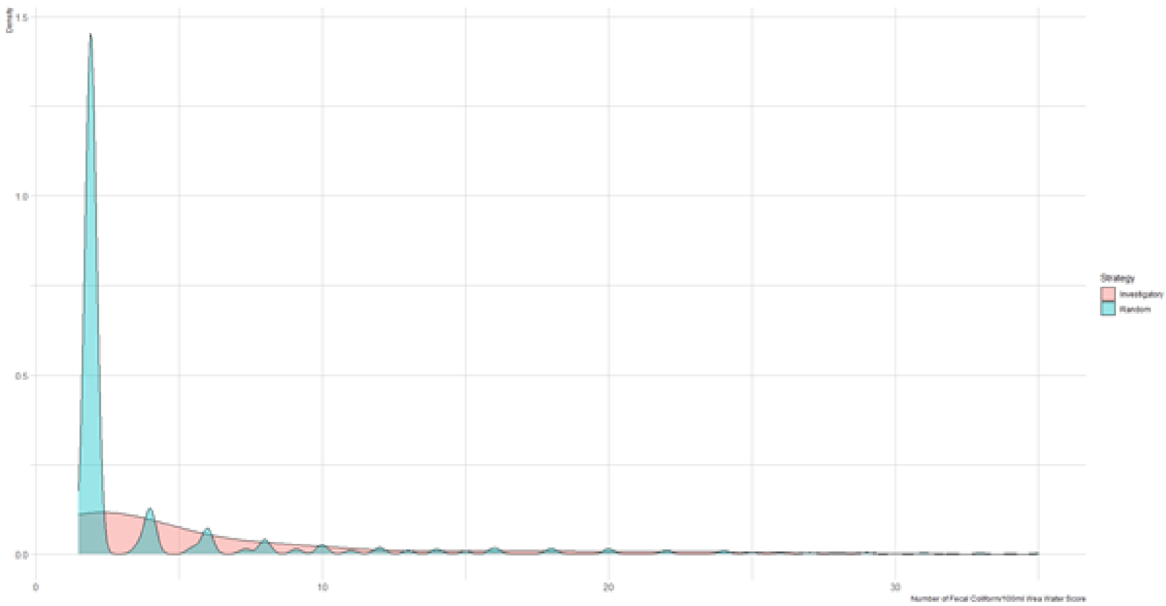


Fig. 3.02. Overall distribution of scores within the entire dataset. While the range of scores is significant, (1.5 to 1,700), most scores are two or below.

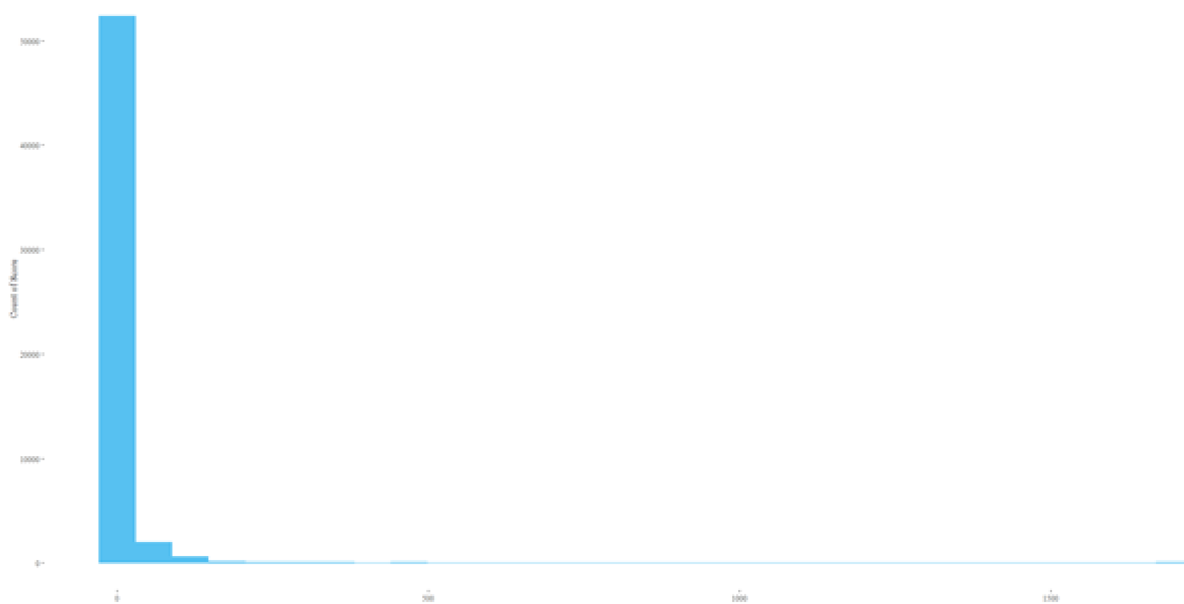


Fig.3.03. Temperature appears to rise and fall seasonally with fecal contamination score.

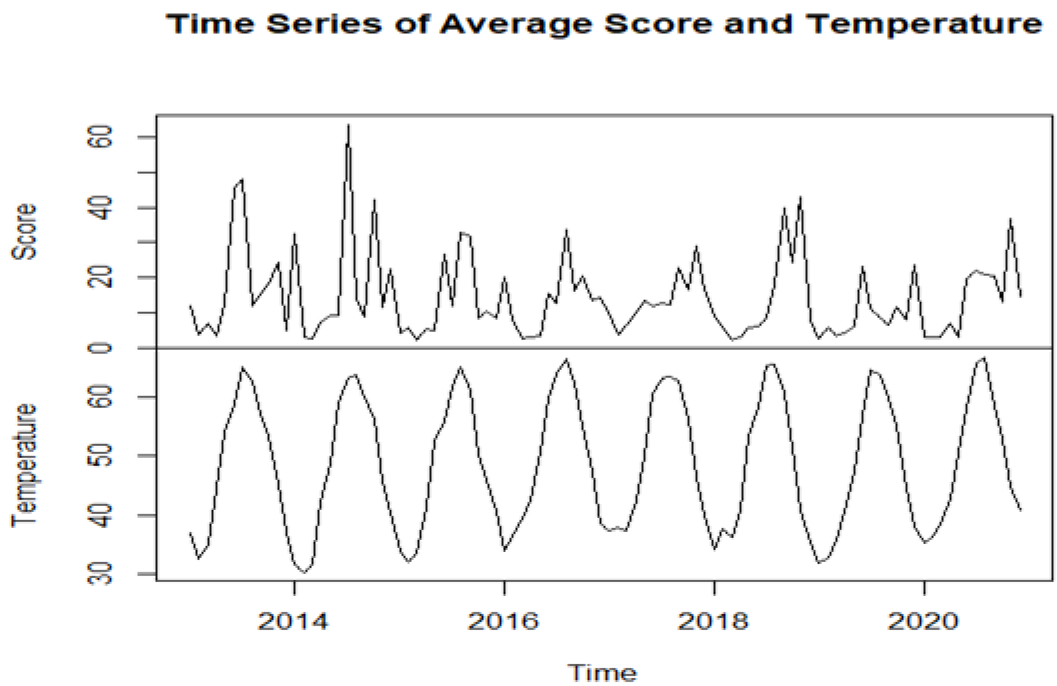


Fig.3.04. Scatterplot shows little correlation between temperature and fecal contamination score. This is further clarified with the R2 score of 0.19.

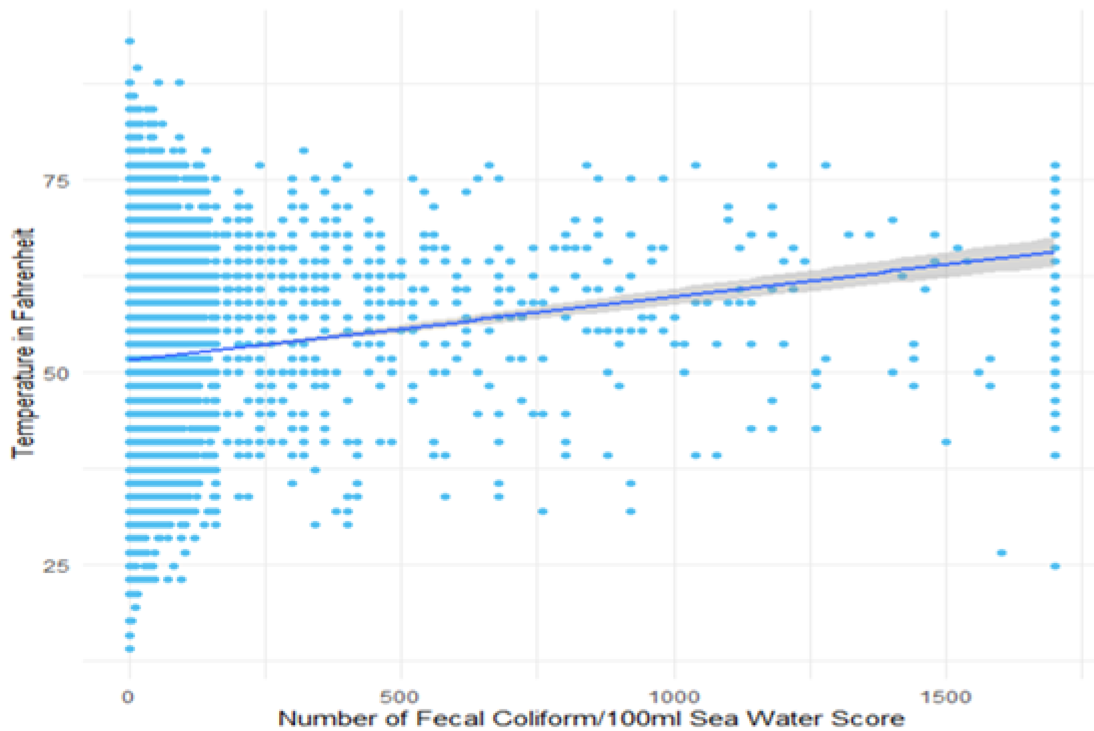


Fig.3.05. Locations of top quartile sites in the trailing four quarters are sporadic and difficult to predict based on past data.

Fig.3.06. Contamination concentrations on Vinalhaven in the 2nd quarter of 2013, the first year Maine began taking samples.

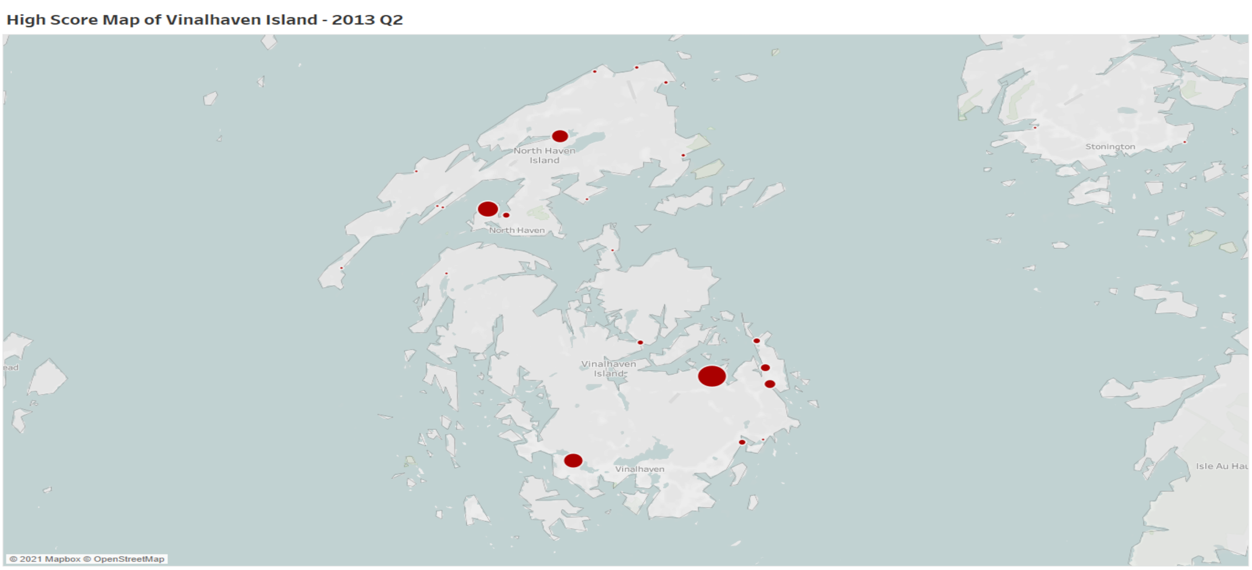


Fig. 3.07. The same image, three years later.

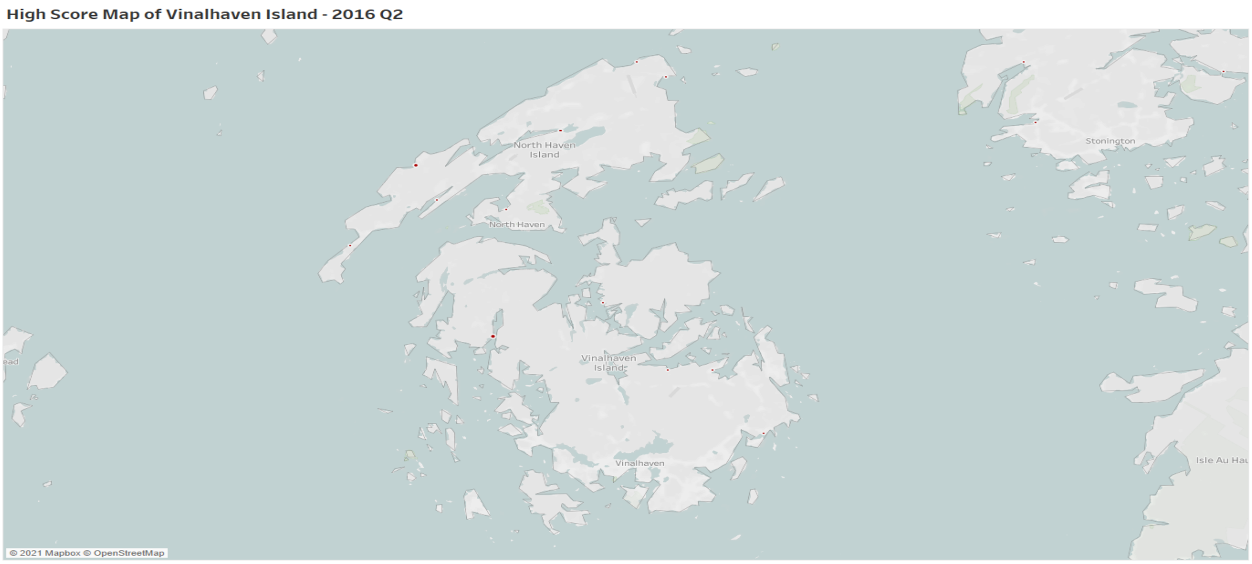


Fig. 3.08.

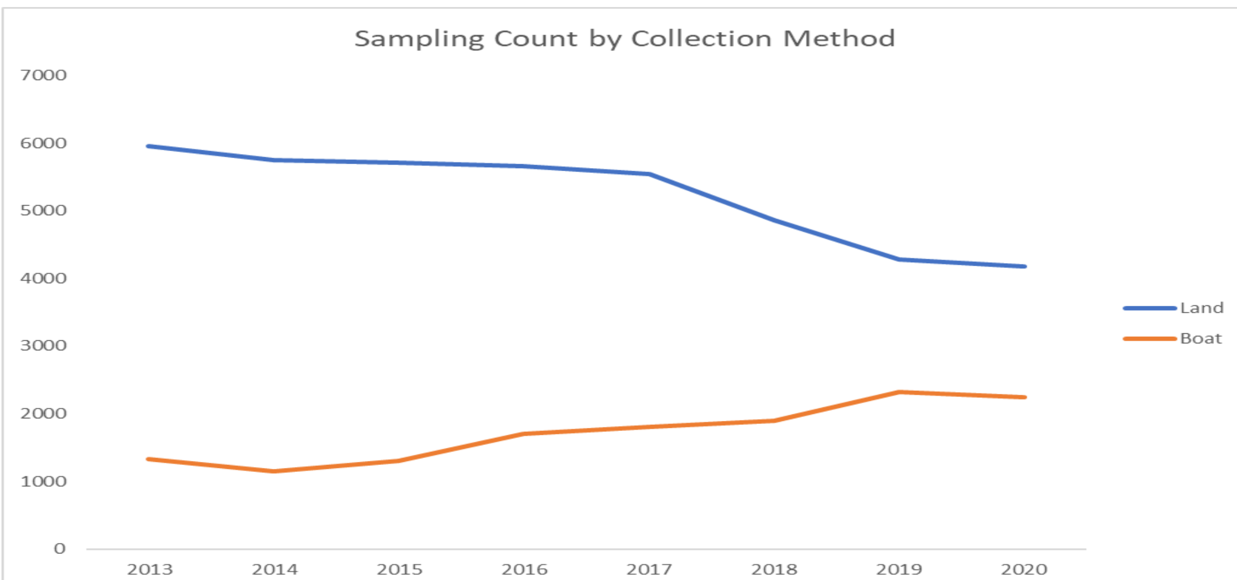


Fig. 3.09.

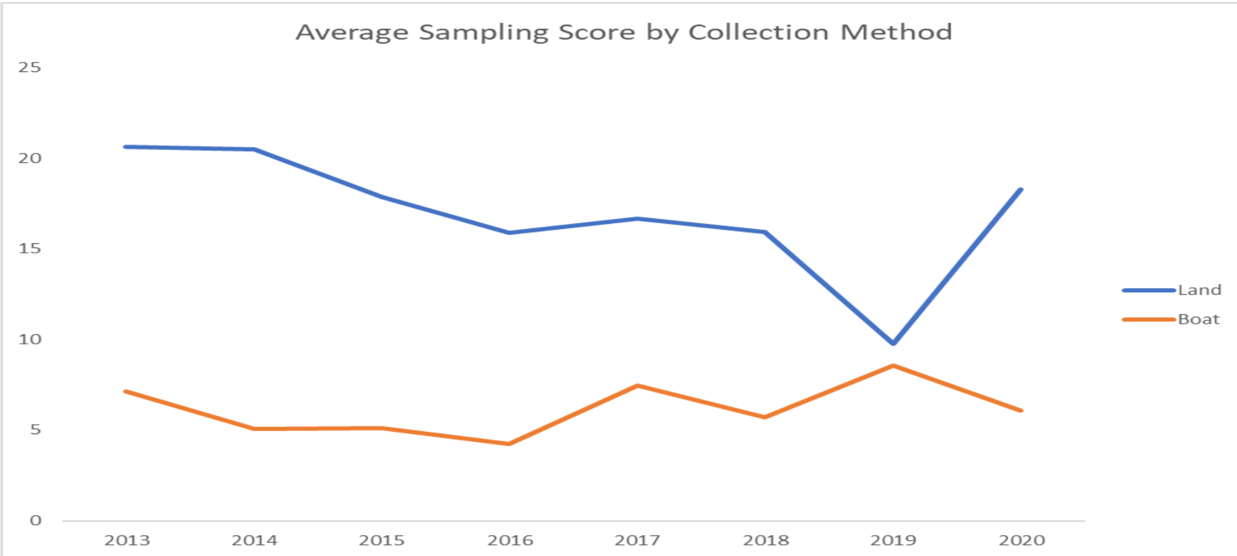


Fig. 3.10. Overall trend shows that as salinity decreases, so does fecal contamination score.

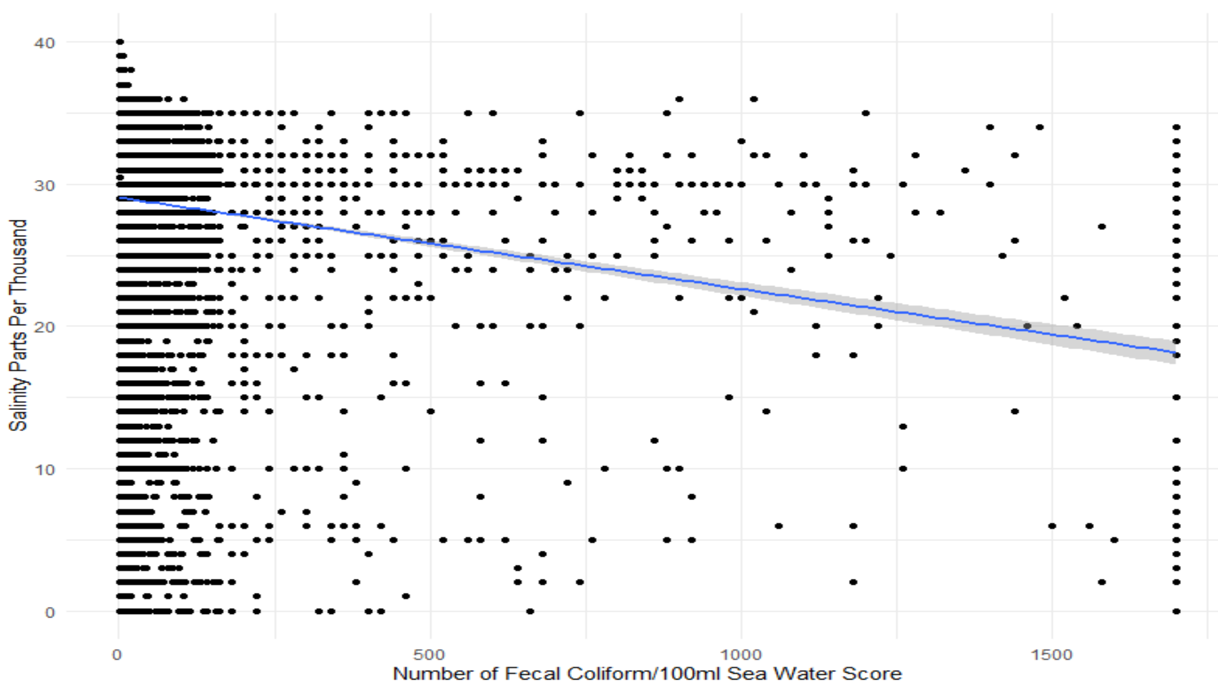


Fig. 3.11. Decision tree identified salinity, adversity (precipitation), temperature and growing area as splits.

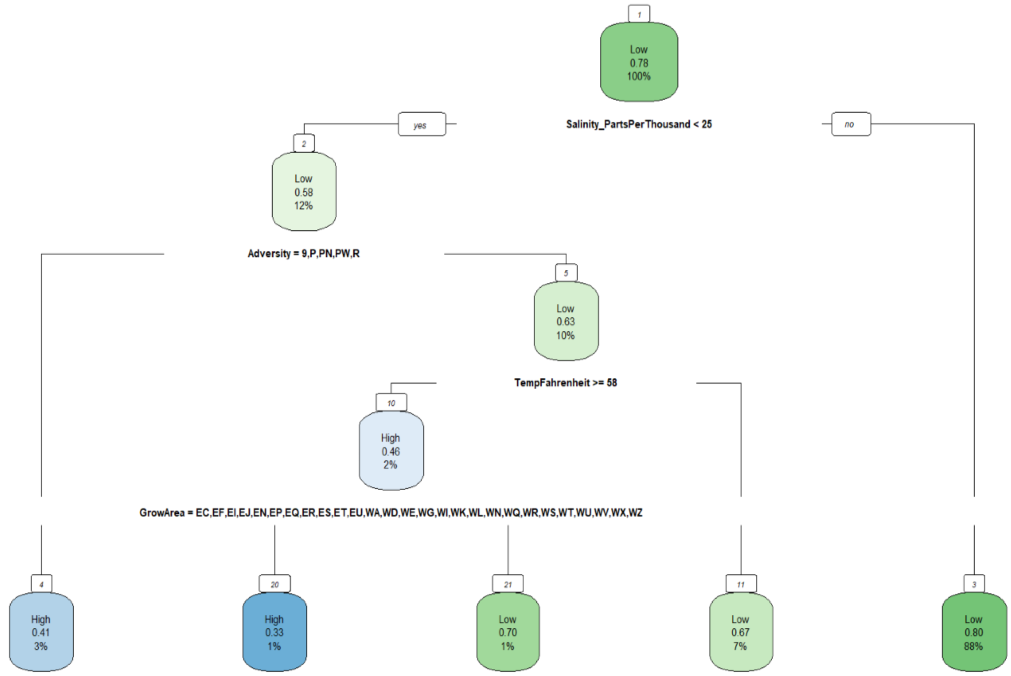


Fig. 3.12.

Confusion Matrix		
	High Predict	**Low Predict**
High Actual	257	2242
Low Actual	167	8396

Bibliography

1. Halmare,M., Joshi,H., Mutreja, S. (2021, January). CubeSat Market Analysis, Size, Share, Forecast by 2027. Allied Market Research. https://www.alliedmarketresearch.com/cubesat-market-A09399#:%7E:text=CubeSat%20Market%20Statistics%202020-2027%20The%20global%20CubeSat%20market,have%20also%20suffered%20due%20to%20the%20COVID-19%20disease
2. Casco Bay Estuary Partnership. (2011). Protecting Shellfish Beds Fact Sheet. Portland, ME: University of Southern Maine, Muskie School of Public Service, Casco Bay Estuary PartnershipCleland, D.T.; https://digitalcommons.usm.maine.edu/cgi/viewcontent.cgi?article=1012&context=cbep-publications
3. Freeouf, J.A.; Keys, J.E.; Nowacki, G.J.; Carpenter, C.A.; and McNab, W.H. 2007. Ecological Subregions: Sections and Subsections for the conterminous United States. Gen. Tech. Report WO-76D [Map on CD-ROM] (A.M. Sloan, cartographer). Washington, DC: U.S. Department of Agriculture, Forest Service, presentation scale 1:3,500,000; colored. https://doi.org/10.2737/WO-GTR-76D
4. Dunbar, B. (n.d.). Space shuttle weather launch commit criteria and ksc end of mission weather landing criteria. NASA. https://www.nasa.gov/centers/kennedy/news/releases/2003/release-20030128.html
5. Emerald Ash Borer (EAB), Agrilus planipennis. Maine Department of Agriculture, Conservation, & Forestry. Retrieved on August 29, 2021 from: https://www.maine.gov/dacf/php/caps/EAB/index.shtml
6. Federal Aviation Administration. Commercial space data. (2021, April 21). https://www.faa.gov/data_research/commercial_space_data/
7. ForeFlight web. (n.d.). https://plan.foreflight.com/map#-69.0683/44.42593/8
8. Frost & Sullivan (July, 2021). Global SatelliteEnabled IoT Growth Opportunities: Disruptive Technologies and Transformative Mega Trends Drive Future Growth Potential and Vertical Market Expansion.
9. Fuller,J., Liao, R., Patel, K., Bandla, S., Jan, M., (2020, July), Building a Global Launch Network: Extending the Reach of Dedicated Small Satellite Launch Using New, Data-Driven Spaceport Assessment Tools. https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=4753&context=smallsat
10. Hartwick, M. A., Urquhart, E. A., Whistler, C. A., Cooper, V. S., Naumova, E. N., & Jones, S. H. (2019). Forecasting Seasonal Vibrio parahaemolyticus Concentrations in New England Shellfish. International Journal of Environmental Research and Public Health, 22, 4341. https://doi.org/10.3390/ijerph16224341
11. Haining, R.P. (2002, November 2) Spatial Sampling. International Encyclopedia of the Social & Behavioral Sciences, 2001, Pages 14822-14827, ISBN 9780080430768, https://doi.org/10.1016/B0-08-043076-7/02510-9
12. Hoque, F.,, Hu, B., Wang, J.,, & Hall, G.B., (2020). Use of geospatial methods to characterize dispersion of the Emerald ash borer in southern Ontario, Canada. Ecological Informatics, 55, 101037. https://doi.org/10.1016/j.ecoinf.2019.101037
13. Larsen, Paul. (2017). Small Satellite Legal Issues. Journal of Air Law and Commerce , 82 (2). p. 275 -306

14. Morin, R. S., Pugh, S. A., Liebhold, A. M., Crocker, S. J., (2015). A regional assessment of emerald ash borer impacts in the Eastern United States: ash mortality and abundance trends in time and space. General Technical Report. https://www.fs.usda.gov/treesearch/pubs/50352
15. MSGC Conducting Study to Determine Market Demand for a Maine Space Entrepreneur and Innovation Complex. Maine Space Grant Consortium. (2018, November 11). https://www.msgc.org/news/msgc-conducting-study-to-determine-market-demand-for-a-maine-space-entrepreneur-and-innovation-complex/
16. NASA. (n.d.). Falcon 9 Launch Weather Criteria. www.nasa.gov. https://www.nasa.gov/pdf/649911main_051612_falcon9_weather_criteria.pdf
17. P90 definition. Law Insider. (n.d.). https://www.lawinsider.com/dictionary/p90
18. Sisson, G. M. A., Jin, Z., Currey, L., Shen, J., & Jia, J.-J. (2007). Developing a Cost-Effective Methodology to Manage Fecal Coliform Loading in Shellfish Harvesting Areas of Upper Chesapeake Bay, Maryland . Estuarine and Coastal Modeling, 543–560.
19. Sheetz, M. (2020, October 2). The space economy has grown to Over $420 billion and IS 'weathering' the current Crisis, report says. CNBC. https://www.cnbc.com/2020/07/30/space-economy-worth-over-420-billion-weathering-covid-crisis-report.html
20. SpaceWorks Enterprises (2019). Williams, C., DelPozzo, S., Doncaster, B. Nano/Microsatellite Market Forecast, 9th Edition
21. Spaventa, S. (2021, May 6). The power of satellite imagery in agriculture & farming. FarmTogether. https://farmtogether.com/learn/blog/the-power-of-satellite-imagery-in-agriculture
22. Maine DMR, Bureau of Public Health, Sanitation & Shellfish Management. (n.d.). The Scoop on Fecal Coliform Pollution https://www.maine.gov/dmr/shellfish-sanitation-management/programs/growingareas/coliform.html
23. Tilburg, C., Jordan, L., Carlson, A., Zeeman, S., & Yund, P., (2015). The effects of precipitation, river discharge, land use and coastal circulation on water quality in coastal Maine. Royal Society Open Science, 2(7), 140429–140429. https://doi.org/10.1098/rsos.140429
24. Torabzadeh, H., Morsdorf, F., Leiterer, R., & Schaepman, M. (2018, June 20). MAPPING forest species composition using IMAGING spectrometry and airborne laser scanner data. Biotechnology, Agronomy and Society and Environment, 22(3), 172-187. https://doi.org/10.5194/isprsarchives-XL-1-W3-437-2013
25. Wallace, R. D. (2020). The New Economic Potential of a Space Complex and New Space Economy in Maine. Dissertation.
26. Weather data Services: Visual Crossing. Weather Data Services | Visual Crossing. (n.d.).https://www.visualcrossing.com/weather/weather-data-services#/editDataDefinition
27. AFP (2020, November 17) Will Small Rockets Finally Lift Off? https://www.digitaljournal.com/world/will-small-rockets-finally-lift-off/article/581273

28. WGME, (2021, August 30) Emergency order further restricts movement of ash tree materials.https://wgme.com/news/local/emergency-order-further-restricts-movement-of-ash-tree-materials
29. WCSH, (February, 2020). Medical marijuana is third in revenue in Maine. https://www.newscentermaine.com/article/money/business/medical-marijuana-is-third-in-revenue-in-maine/97-582e2c65-30be-4212-a76d-f80d5e1d6514
30. USDA Forest Service (2020) Forests of Maine, 2019. Resource Update FS-236. Madison, WI: U.S. Department of Agriculture, Forest Service. 2p. https://doi.org/10.2737/FS-RU-236