MIC – Smart Beach:

Wave and Beach Security Analysis

CRYdata Solutions

Authored by: Cassandra Forlani | Ruwindhu Dilanga Chandraratne Hettige Don | Yi-Chen Hsiao

December 5th, 2022

Georgian

NUCLEAR INNOVATION INSTITUTE™

Table of Contents

Introduction

Beaches around Lake Huron are popular tourist destinations. However, there are some potential risks such as quick underwater rip currents. Local community organizations are seeking solutions to educate and maintain beach safety.

This project is part of the three-year Smart Beach project administered by the Municipal Innovation Council (MIC). It aims to use innovative technologies to improve beach safety. Launched at Station Beach in Kincardine on May 25, 2022, the research team of the Smart Beach project has collected weather and wave data using a RAEON (Real-time Aquatic Ecosystem Observation Network) buoy in Lake Huron from Spring to Autumn in 2022.

This project aims to utilize the collected buoy data and the publicly available Environmental Canada and NOAA data to examine if the inshore water movement (wave conditions and currents) and drowning incidents can be predicted by the offshore buoy data and weather conditions. The results of machine learning models are expected to support the Smart Beach project further to develop a real-time beach security information system for beachgoers and improve safety at Kincardine's Station Beach.

Project Description

Objective

- Examine if the inshore water movement and drowning incidents can be predicted by the offshore buoy data and weather conditions to support the Smart Beach project's next steps for improving beach safety.
- If inshore water movement and drowning incidents can be predicted by predictive models, identify the best predictive modeling to support the Smart Beach project's work in developing the real-time beach safety information system and education programs.

Requirements

- Must acquire the following data in order to perform predict modeling: (1) inshore water movement (wave, current) and weather/wave data from NDBC buoys in Lake Huron collected by the Smart Beach project research team; (2) historical weather and wave data from NOAA and Environment Canada that match the location and period data collected by the Smart Beach project research team; (3) drowning and rescue history records at Kincardine and along the Huron shore of Bruce County.
- Must have at least one team member proficient at machine learning/predict modeling and coding to perform predictive analysis.
- Need to manage project timeline and ensure project output progress meets client's need.

In and Out Scope of Project

In scope

- Examine if inshore water movements can be predicted by the weather/wave data from NDBC buoys, NOAA and Environment Canada.

- If inshore water movement can be predicted by predictive models, identify the best predictive models.
- Utilize the above model to predict the wave, current and temperature conditions for past drownings and rescues at Kincardine and along the Huron shore of Bruce County.

Out scope

- Deploying the predictive model on the client's website.
- Build a real-time interactive dashboard.
- Make educational content for beach safety.
- Conduct in-field observations.

Outcomes and Benefits

Expected outcome

- A summary about the predict modeling effectiveness evaluation
- Predictive models regarding inshore water movement.
- Predictive models regarding past drownings and rescues at Kincardine and along the Huron shore of Bruce County.
- The recommended predictive models.

Expected benefits

- If the predictive modeling is effective, analysis result would be beneficial to the Smart Beach project's next steps.
- The beach safety for beaches around Lake Huron is expected to increase when the Smart Beach team utilizes the predictive models to develop real-time beach security information system and education programs.

Deliverables

- Finalized report.
- The predictive model.
- Python program code and scripts.

Team Profiles

Name	Main Roles	Education	Experience/skills
Cassandra Forlani	Team lead and Communicator	Bachelor of Mathematics (Honors Mathematical Studies major with a Pure math and Psychology minor) Bachelor of Education	Academic settings: -Some statistical knowledge through stat courses -Collaboration and problem-solving skills when working on proofs of theorems Teacher setting: -Prepare/lesson plan/organize classes that range from 8 weeks to 6 months. -Dealt with 30 to 150 students at the same time in a virtual/hybrid/in-class setting -Collaboration, mentorship and leadership skills -Knowledge and comfortability with google classroom, Microsoft teams, and zoom

			Tools: Apache-Hive, Power BI, Python, R, Report Builder, and SSMS Data Experience: Publicly available datasets via Nasa Earth Data, Kaggle, and Statistics Canada
Ruwindhu Dilanga Chandraratne Hettige Don	Developer	Post Graduate Certificate in Artificial Intelligence, Implementation and Design Bachelor of Engineering in Software Engineering	Academic setting: Optimum Predictive Model Identification for Customer Churn in Music Streaming Industry Body Measurement Extraction using Open Pose Business Setting : Worked with Constraint Based Scheduling Software for Manufacturing Worked in a Proof of Concept project for Next Gen Manufacturing Order Planning using Esri ArcGIS Implemented IFS (https://www.ifs.com/) Manufacturing Visualizer to a new Web based Client Worked with converting C# IFS Access Provider into OData Provider for IFS Manufacturing Scheduling Client. Tools: C#, C++, Python, R, Power BI, Apache-Hive, Oracle SQL, MySQL, Django, Jira, Esri GIS Data / ML Experience: Mostly publicly available datasets on Kaggle such as the KKBOX Dataset and Alcohol Sales in Iowa
Yi-Chen Hsiao	Analyst	PhD in social science field (Health and social Welfare).	Academic settings: Predictive model of health behaviors Business settings: - Work as Data Analyst for 1 year - Supported sales and marketing teams by doing market research, business analytics, back-end data analysis - Entrepreneurial & working with client experience: Helped small business owners to form branding and marketing strategies via market research (quantitative & qualitative) and data analytics (descriptive & predictive analysis) - Co-founder of a female entrepreneurship association in Taiwan (https://herattitude.org/). Facilitate experience and resource exchange between members. Speaker of 8 lectures about data-driven decision making. Tools: Python, R, Power BI, Apache-Hive, SPSS Dataset experience: Small pilot survey, health survey data (sample size > 18,000), datasets derived from CRM (e.g., e-commerce platform, travel agency -order number > 500,000)

Execution

During the initiation phase of the project, the research team from Municipal Innovation Council agreed to provide the data gathered at Kincardine beach during the summer. The data would include wind velocity, wind direction, air temperature, air pressure and other variables collected using the RAEON buoy placed near Kincardine beach. In addition to this, the project team would collect data from publicly available historical data from the NOAA (National Oceanic and Atmospheric Administration) and Environment Canada.

The dataset then would be cleaned via the use of Python Scripts as the data is recorded directly from the RAEON buoy. Next the team would perform both statistical analysis and EDA (Exploratory Data Analysis) on the dataset. The analysis would support the team in gaining more insights on the dataset. This analysis would include performing a time series analysis on different variables in the dataset. The Time series analysis would be used to identify which hours of the days have extreme conditions at Lake Huron.

Next the team would be performing many feature engineering techniques to prepare the dataset in creating a machine learning prediction model. Feature Engineering entails feature scaling, feature categorical encoding and cleaning outliers. The dataset would then be spilt into 2 sets, one set for training the machine learning model and the other set for testing the machine learning model.

An iterative process would be used to select the ideal machine learning algorithm to create the predictive model. Initially a simple regression would be selected, and the training set of data would be used to train a simple machine learning model. When training the machine, learning different combinations of hyperparameters would be used to create the best machine learning model. Next, the created machine learning model would be tested on the “test set” of the dataset. Using the historic data collected from the public data sources, the model would be validated by comparing the wave properties. Finally, the model would be packaged and provided to the MIC team. If the model trained does not produce accurate predictions, a new model would be created, and the process iteration would be initiated.

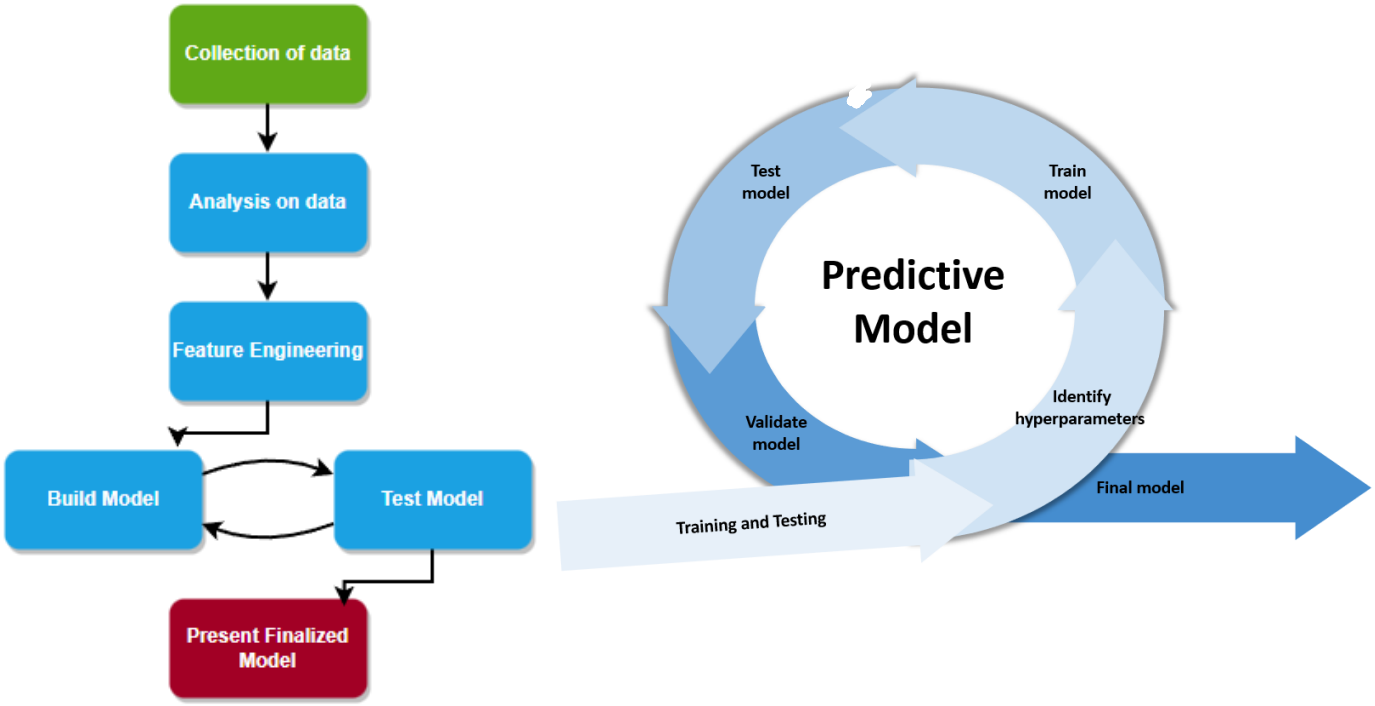
Techniques

Techniques	Use of Technique	Where is the technique used	Why was this Technique Selected
Predictive Regression Modelling	The technique would be used to create machine learning model for predicting wave conditions.	To predict inshore significant wave height, wave period, wave direction, current speed, and temperature profile	A regression model would be suitable for this project, since the project is focused on predicting the trend of wave conditions.
Data Cleansing	The data obtained from public sources and MIC would need to be cleaned.	Clean the data and identify which features are appropriate to be used as input for the prediction model.	The collected data from the RAEON buoy, and other public sources may contain empty values and other anomalies. Thus, these values need to be cleansed.
Feature Engineering	Used to scale and identify the desired variables to be predictors for the machine learning model	Feature engineering would be used to prepare the dataset prior to creating a machine learning model.	Prior to training a machine learning model, the variables need to be scaled and manipulated so that the machine learning model is more accurate.
Agile development methodology	This methodology would be used as the development strategy for the project. With this, the client would be involved from the beginning and kept up to date.	This methodology would be used from the design stage of the project. MIC will be able to input their feedback after each iteration. This will allow collaboration and the ability to refine the solutions.	This methodology is suited for this implementation (vs a traditional waterfall development) as MIC can give their input on the implementation during the development stage.

Tools

Tool Name	Use of Tool	Where is the tool used	Why was this Tool Selected
SharePoint	Project management / Process Tracking	SharePoint would be used throughout the entire project as it allows the team to track their progress in the implementation.	SharePoint is an ideal tool for collaboration as it allows users to share documents and track progress of the project.
Microsoft Teams	Communications with stakeholders	MS Teams would be used as the primary communication tool; it would be used during meetings with MIC as well as internal team meetings	Microsoft Teams is widely used in all fields of work. Therefore, it would be the most appropriate tool to be used for communication.
Scikit Learn	Predictive Modelling and Data Preprocessing.	Scikit-Learn is a well know python library which will be used in the predictive modelling implementation. In addition to this, it would be used to feature scaling and feature extraction.	Scikit-Learn is one of the best python libraries when it comes to generating regression models. Furthermore, its feature engineering capabilities are unmatched with its competitors i.e. feature hashing, feature extraction and feature importance to name a few.
Python Data Analysis Library (Pandas)	Data Cleaning and Data Analysis	Pandas would be used to clean the dataset and remove outliers present in the dataset. In addition to this, Pandas would be used to create simple plots which can be used to perform analysis on the dataset.	Panda is widely used in the industry to clean and perform simply analysis on datasets. This is due to Pandas data frame allowing effective data manipulation.

Development Process Visualization



Potential Challenges

Even though some challenges may develop, we plan to provide the Smart Beach project with completed analyses/report that answers key questions. Still, these potential challenges may impact our process and schedule:

- Incomplete/inaccurate data from NDBC buoys which can affect the ability to accurately predict specific variables[1]
- Finding insufficient/precise data to do a prediction model solely with public data from Environment Canada and NOAA
- The prediction model not being able to identify with accuracy specific variables[1]
- The Smart Beach project needs changed
- Scheduling conflicts which can defer input meetings

Note[1]: variables = wave height, wave period, wave direction, current speed, and/or temperature

Risk and Issues Management

When confronted with issues, the Team will determine the best way to deal with these inconsistencies. If an issue creates obstacles in completing an objective, the Team will contact the Client immediately to notify them, make recommendations and act according to their instructions. Dedication by the Team to resolve issues may impact their ability to achieve the desired timeframe's anticipated results.

Timeframe

The Team will exercise its best efforts to respect deadlines and deliver the anticipated objectives.

Project Team

We are confident in our Team but realize that other demands may come up at certain times (e.g., exams, illness, personal events). If the Project's progress is delayed at any time, the Client will be notified immediately and advised of the Team's plan to accomplish the Project's requirements.

Model Selection

Model selection may be time-consuming. If a misfit is determined, there may not be enough time to select a new model and to complete the Process. The Team will exercise its best efforts to ensure the model selection is appropriate for this Project.

Communications

The Client is committed to supporting the Project Team in achieving the objective and providing additional resources if required.

Scope Creep and Changes

Scope creep can occur when a project's scope is not adequately defined, documented, or controlled. Also, when added into the Project, either by the Client or the Team itself. To avert a project threat, the Team will clearly and accurately document each requirement while planning and only apply their resources to achieving those requirements. When scope changes develop, the Team will document the requirements, identify the impact of the change on the Project, notify the Client, make recommendations and act according to the Client's instructions.

Code of Conduct

Ethical Considerations

- We acknowledge our strict duty of confidentiality.
- We will ensure the observance of ethical practices demonstrating respect, honesty and dignity.
- We will thoroughly follow the Client's directives without deviation from the objectives.
- Value is always present in our thinking.
- Validated data will solely form the basis of our observations, conclusions, recommendations and decision-making.
- We undertake to use holistic analytics strategies and repeatable processes.
- We will thoroughly explain and document our analysis in detail for the Client. We will be transparent in all our activities.
- Moreover, we will check and recheck our research for quality, accuracy, completeness and integrity before presentation to maintain the validity and credibility of the results.

- We will proactively circumvent unethical behaviour, such as exaggerating the results of our research.
- We will not falsely interpret, fabricate, embellish or otherwise misrepresent the data to validate our findings or change or omit details favoring making an analysis fit a hypothesis.
- We will be forthright and accountable if we make mistakes.
- We will tell the truth, even if it is bad news.
- Failure to observe deadlines is considered a breach of ethics.
- We have no conflict of interest and will not benefit from this Project.

Information Management

- We acknowledge that the data is an asset of any organization. The consequences and repercussions of unethical conduct when dealing with an organization's data can be significant and affect an organization's reputation, relationships and, ultimately, its revenues. Even the perception of unethical data handling has the power to undermine both internal and external trust.
- We give assurance that all the information provided is in our trust. It will remain private and protected from damage or alteration unless authorized.
- We will sign a Non-Disclosure Agreement that is acceptable to the Client and adhere to its conditions.
- We will notify the Client of the discovery of any sensitive information at any stage, namely personal identifying information or confidential information. We will make recommendations for addressing the issues and act according to their instructions.
- Original data and other information entrusted to us is stored in a secure location, such as SharePoint, and will remain unaltered.
- Copies of original data will be used for cleaning, discovery, manipulation and analysis.
- Naming conventions and version controls will segregate documentation, datasets and work product. Each will align with the applicable phase of the Project.
- All work will only be conducted through secure applications.
- All our workstations are protected by a password that is unknown to anyone other than the assigned user. No one has access to our workstations, including colleagues, family members and friends.
- All information sharing is secured from potential risks using encrypted channels, such as SharePoint.
- Discussions about the data, analysis and observations will only be amongst client-authorized collaborators and our course instructors.
- All information entrusted to us will only be used for its intended purposes unless specifically approved by the Client.

Milestones

Milestone	Complete By
Acceptance of Project Charter	December 7th, 2022
Requirements Phase	January 12th, 2023
Data Cleaning and Analysis Phase	January 26th, 2023
Development and Testing Phase	March 23rd, 2023
Release Phase	March 30th, 2023
Client Presentation	April 3rd, 2023
Project Close-Out Phase	April 5th, 2023

Conclusion

By April 5, 2023, we will be providing an overall analysis that allows you to predict inshore significant wave height, wave period, wave direction, current speed, and temperature profile. We will be doing this by first processing the data collected from you and the Environmental Canada/NOAA data.

If time permits, we will also be creating a prediction model on the following:

1. The wave, current and temperature conditions for past drownings and rescues at Kincardine/Huron shore of Bruce County.
2. The number of people in the water and on the beach based on the wave conditions.

Our team is delighted to participate in this opportunity. We are confident and committed to successfully completing this project and provide the best outcome based on our collaboration with you.

We look forward to working with you these upcoming months.

Thank you,

Georgian College Big Data Analytics Team – CRYdata Solutions

Cassandra Forlani | Ruwindhu Dilanga Chandraratne Hettige Don | Yi-Chen Hsiao