Predicting the likelihood of the need to launch a RNLI rescue boat in Ireland based on the Weather and Bank Holidays

MSc Research Project
MSc in Data Analytics

Elmarie Fryer
Student ID: 20220278

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Elmarie Fryer
Student ID:	20220278
Programme:	MSc in Data Analytics **Year:** 1
Module:	Project Report
Supervisor:	Jorge Basilio
Submission Due Date:	23rd August 2021
Project Title:	Predicting the likelihood of the need to launch a RNLI rescue boat in Ireland based on the Weather and Bank holidays
Word Count:	6309 **Page Count** 20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Elmarie Fryer ………
Date: 23rd August 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predicting the likelihood of the need to launch a RNLI rescue boat in Ireland based on the Weather and Bank holidays

Elmarie Fryer
Student ID: 20220278

Abstract

This paper investigates the relationship between the need to launch a lifeboat and the weather. The ability to predict when a rescue boat will have to be launch will greatly assist organisations, such as the RNLI, that are dependent on volunteers for their efforts. Being able to predict when a launch is most likely will assist volunteers in planning personal activities, and add to the local, institutional knowledge currently used by volunteers. The impact of the volume of water users, as represented by Bank Holiday weekends, is also considered.

Data used include Irish weather data from Met Eireann, Incident data from the RNLI, Irish Bank Holidays, Seasonality, and a Calendar and the architecture used include an Azure SQL Server and RapidMiner hosted on an Azure Virtual Machine. The project was executed on Cloud based technologies to reflect the reality of geographically dispersed volunteers.

Nine predictive models are evaluated in terms of Accuracy, Class Prediction and the Cost to Compute. Gradient Boosted Trees and Deep Learning are found to be the best fit/ models deployed. Further research is also identified taking gender, skill level and the types of incidents into account. This will inform policy making for water safety.

1 Introduction

There are on average 25 accidental coastal drownings in the Republic of Ireland (Kervick et al., 2016) each year.

Speed is critical during a rescue situation (Shih et al., 2018) and survival becomes extremely unlikely after 30 to 90 minutes depending on the water temperature (Tipton & Golden, 2011). The RNLI measure time to casualty from the time that they are informed until the casualty is reached. That includes time to assemble as small craft survivability drops from almost 100% up to 25 minutes to approximately 20% after 55 minutes. (Pitman et al., 2019). Studies into CPR for reasons other than drowning have found that CPR resuscitations on lifeboats occur under mainly hostile conditions (Seesink et al., 2019) and that only 30% of patients responded to CPR and in survival rates vary from an average of 8% to 2% or less if the delay is longer than ten minutes (Capucci et al., 2002).

The Royal National Lifeboat Institution (RNLI) defines itself as the "charity that saves lives at sea"[1]. And provides rescue from 238 lifeboat stations across Ireland and the UK. Of the almost six thousand crew 95% are volunteers (Royal National Lifeboat Institution, 2021). There are also 54 Community Rescue Boats Ireland (CRBI) that not affiliated with the RNLI, and trained by water safety Ireland. (Water Safety Ireland, 2021). Of these 40 are a declare resource used by the Irish Coast Guard along with the 35 RNLI Stations. Many of these groups were set-up in communities after a tragedy as there was no other resource nearby.

Rescue attempts are mainly reliant on thousands of unpaid volunteers (Macleod, 2018), who have to get to the lifeboat station[2] from wherever they are once they have been paged. Managing unpaid volunteers can be very challenging (Munro, 2011; Ward & Greene, 2018) and in the case of the RNLI studies (O'Toole & Grey, 2015) have found that many volunteer crew at the RNLI have familial and community ties, share knowledge only among those they trust and have a share experience with (O'Toole & Calvard, 2019; Roger, 2004). Expert knowledge is most often related to first-hand experience in local waters and a significant percentage of volunteers have family members who are (or were) volunteers (Grey & O'Toole, 2020).

As the demographics of water users and the volunteers change there is less local knowledge. This is where data analytics and machine learning can be a useful tool. The ability to predict whether there's likely to be a requirement for rescue can allow volunteers to plan their activities. The focus of the research objectives and research questions is on volunteer activity (i.e. launching the boats). The reason for, and outcome of, the launch is therefore ignored for the purposes of this project, and should be investigated further in future work.

The objective of this paper is to describe the weather conditions prevalent during a launch and ultimately predict the likelihood of a launch given the weather.
To that end the following research questions are identified:
Can we predict the likelihood of the need to launch an Irish rescue boat based on the weather?
Can we improve the prediction by adding in additional variables to estimate number of people?

This paper will contribute to scientific literature by using Irish data to evaluate the relationship between the weather and incidents at sea, and by extending work related to weather and traffic accidents to marine accidents.

[1] https://rnli.org/about-us
[2] https://rnli.org/what-we-do/lifeboats-and-stations/what-it-takes-to-launch-a-lifeboat

The report has the following sections:

- An analysis of the work related to rescue in the marine environment, and an analysis of work combining weather data with incidents in Aviation and Traffic.
- A discussion of the research methodology, including the process followed and the sources and transformation of the raw data
- The Design Specification including tools and resources used
- The implementation of the design including results and suggested models
- Evaluation of the models created during the implementation and selection of the models with the best fit
- Summarising the results of the paper and suggestions for future research.

2 Related Work

2.1 Marine Rescue

Pitman, et al (2019) investigated the causes of sea rescues by RNLI using the data from six years (2011-2016) of rescues in UK waters. The factors used in their study included environmental conditions (visibility, sea state, etc.), lifejacket wear, and response times for rescue. They used traditional statistics and developed a Poisson model to evaluate their data.
The main finding was that wearing a life jacket significantly improves survival rates and this has led to an increased focus in the RNLI on campaigning the use of life jackets.
The results show a clear seasonal impact with incidents during the winter resulting in a higher proportion of fatalities. Users fishing from shore are also shown to be at high risk of drowning and the study suggested policy improvements across sea safety in the UK.

A limitation of this study was that the data used was solely limited to the RNLI database and the sea and weather conditions recorded there. This isn't always up to date or complete.
The classifications for weather were also very broad including several values with no impact on the results. This has been shown in other literature to be a weakness and that less categories could lead to stronger results. In addition, this study only used the data related to the UK.
As seasonality has an impact on survival, seasons were added to my dataset. I also combined the RNLI data with external weather data to improve the results. The added benefit of this is that there were results available for days without any incidents which will allow me to determine whether or not using weather data will allow the prediction of when incidents may occur.

There were three related studies that focused on the risk of drowning at beaches:

- Stokes, et al (2017) used multiple linear regression (MLR) and a Bayesian belief network (BBN) to select the beaches most in need of life guard services in the UK. The main factors determining the selection was risk of life due to morphology type of the beach, car parks, and headlands. BBN was found to be superior at allowing for variables based on subjective expert opinions but had a lower accuracy that traditional MLR. Limitations of the study was the use of only a limited dataset (3 years of data for a small selection of beaches under consideration) and the use of only one model. To improve on the study use more data and additional models.

- In Australia researchers found that experience improves the ability to swim well in surf (Tipton et al., 2008) by testing the speed with which various swimmers completed set distances. They found that more experiences swimmers fare better in various conditions. A limitation of the study was that the participants were all involved in life saving activities and therefore the study excluded inexperienced, casual swimmers.
- A further study (Morgan & Ozanne-Smith, 2013) found that frequency of rescues and the outcome thereof was influenced by offshore winds and higher temperatures and also by the experience and health of the swimmer. Limitations of the study were the judgement of the lifeguard used to determine experience, and the lack of explanation for the higher numbers of people on the beach.

A study of 160 near-drowning experiences in Finland found that there is often an “absence of rational thinking “ in a drowning situation and that not being alone is often enough to make a difference (Toivonen, 2016). The experience of a swimmer may therefore not always be a factor when it comes to an emergency at sea. There is no study that includes sailing experience in Leisure activities. A weakness not addressed by this paper.

Studies related to commercial shipping include:

- A study of marine accidents using the Swedish Maritime Administration database clustered the data into eleven categories (Mullai & Paulsson, 2011). Fields with large numbers of values were reduced to smaller groups using judgement. This is a limitation that can be improved by using Text Analytics.
- A project using risk modelling of the safety factors involved in marine transportation accidents using Markov and Markov Chain Monte Carlo simulations (Faghih-Roohi et al., 2014). This is an example with limited data and the models in this paper are not suitable for a large dataset such as the RNLI dataset.
- (Eliopoulou et al., 2016) found that fishing vessels in European waters have aged significantly since 2006 and are therefore more at risk of engine issues requiring rescue.
- A Nigerian study reviewed the causes of marine accidents in terms of the impact of human factors. (Oluseye & Ogunseye, 2016). Data was collected from marine operators using surveys and analysed using traditional statistical methods. Human errors were caused by substance abuse, fatigue, and stress leading to recommendations for regulation.
- Viellechner & Spindler (2020) analysed the impact of weather on container shipping delays using 10 regression models and 7 classification model. They found that Neural networks and SVM gave the most accurate results. A weakness of their study was that the time and cost involved in each of the methods was not included. This means that there’s no discussion about the practicality of deploying the various methods at scale.

Advances in Technology that may impact future work:

- Technology to assist in tracking a target at sea using a camera and various machine learning techniques are constantly evolving (Bibby & Reid, 2005) (Xinqiang et al., 2017). These advances may provide the ability to use drones in future for some of the activity currently requiring the launch of lifeboats and could positively impact on the number of hoax calls that waste a lot of time.
- The use of buoys to obtain offshore wind measurements compared to the European Centre for Medium Range Weather Forecasts (ECMWF) of wind and wave models found that the forecasted models are continuously evolving and becoming more accurate .(Dee et al., 2011; Gallagher et al., 2016) This means that, in future, offshore data can be added to the onshore data used in this study.

2.2 Machine Learning Models in similar datasets (the impact of weather on traffic and aviation

Papers related to Aviation:

- As a follow up to a paper stating that there was a "killing zone" (Craig, 2001), where pilots with hours in excess of a certain number were more likely to crash using linear regression, the authors found that by using non-linear models a different picture emerges (Knecht, 2013) that has a more accurate prediction of when a pilot is likely to crash.
- A project with a variety of machine learning techniques to predict conditions that increase the likelihood of aviation accidents (Burnett & Dong Si, 2017) and comparing the results to prior work using only statistical techniques. The data used included old data and the authors noted how badly a model can be affected by not cleaning bad data. Decision Trees and other models taking account of subjective measures of the "human factor" were the least accurate.
- A study into the prediction of serious flight incidents based on a large number of variables (Ni et al., 2019) found that a complex model using a deep belief network with principal component analysis gave the most accurate results. It is an extremely complex model and very difficult to explain and therefore not suitable for a model to be used by volunteers.

Papers related to Road usage and accidents that involve weather:

- Analysis of severity of traffic accidents found that models struggle to predict small values. (Li et al., 2012). This is due to the small number of accidents. The authors did not attempt alternative methods and none of their models were accurate. Future research was suggested, and the limitations were addressed in recent papers.
- Modelling to predict the cause of road traffic accidents found SVM gave the most accurate results (75%) (Mohamed, 2014) assisting authorities in determining causes and developing strategies for prevention. The time required to process these models was not provided.

- Recent work include: (1) Weather related crash prediction in Connecticut (Mondal et al., 2020) that compared different models and found Random Forest the most accurate at 73%. (2) Generalised additive modelling and Random Forest to analyse the impact of weather of traffic accidents in Mexico City (Bailey et al., 2020). Found that there wasn't as large a difference between the results of the two models as expected. (3) A study of the impact of weather on road surface conditions (Kim et al., 2021) used artificial neural networks (ANNs) had an accuracy in excess of 78% and found that there are significant benefits to road users to forecast when the roads conditions are more likely to lead to an accident.

A study on the impact of weather impact on traffic (Harikrishnan & Postwala, 2021) compared several models and found low accuracy. They didn't group or cluster their variables and future recommendations include clustering to improve the accuracy of the results.

In the study of weather and freeway crashes in China (Zeng et al., 2020) the complex SVM was inconclusive. The authors used estimated weather and suggested improvements include the use of actual weather data.

A study of the impact of weather on transportation performance found that more accurate results are obtained by limiting classifications to small group (Zou et al., 2020). E.g., classifying weather from various categories into Normal and Adverse gave more accurate insights into the data.

Improvements to prior research predicting the impact on traffic caused by road accidents based on prevalent weather conditions (Harikrishnan & Postwala, 2021) by using more sophisticated models (Decision Trees with Classifiers and voting Ensembles, artificial neural network (ANN), and Gaussian Naïve Bayes. These more models give greater accuracy but require a more computing power. In a large city this cost is justified by the time saved after rerouting traffic. A limitation is that this is not suitable for smaller projects.

2.3 Summary

There is a gap in the literature as none of the studies include Irish data – this paper will address that gap.
Items identified from the related work to include in my design include:

- Data cleansing improves the accuracy of models (Burnett & Dong Si, 2017)
- Not grouping or clustering source data (Harikrishnan & Postwala, 2021)
- Using estimated, not actual, weather data (Zeng et al., 2020)
- Keeping the potential values in a category as small as possible (Pitman et al., 2019; Zou et al., 2020).
- Include computing power/time required in the evaluation of models (Harikrishnan & Postwala, 2021; Viellechner & Spinler, 2020)

3 Research Methodology

The CRISP-DM (CRoss Industry Standard Process for Data Mining) Process model will be used as it is technology independent. (Wirth & Hipp, 2000) (Chapman et al., 2000).
This model is discussed in more detail in section 4.

Per the CRISP DM Model, the methodology followed is discussed in terms on Data (Collecting and Preparing), Modelling, and Evaluation.

3.1 RNLI Data

Table 1: RNLI Returns of Service per Country[3]

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
GBR	7594	8474	7962	8140	7623	7500	7555	7419	7979	7365	7819	8997	7757
IRL	672	700	722	707	681	819	806	817	852	789	699	841	725
n/a	21	38	33	27	52	12	19	14	5	4	2	1	5
Total	**8287**	**9212**	**8717**	**8874**	**8356**	**8331**	**8380**	**8250**	**8836**	**8158**	**8520**	**9839**	**8487**

Available RNLI Data related to Returns of Service from the RNLI Open Data project. Due to the impact of Covid 2020 was excluded(De Vos, 2020). The UK was later excluded to focus on Ireland and data was limited to 2015 to 2019 to improve the model speed for deployment. A shorter period is consistent with the size of studies in the relevant data reviewed.

Table 2: Abridged Irish Returns of Service from 2015 to 2019 [4]

Sea Conditions At Incident	Waterside	Personal Water Craft	Boat	Other	Grand Total
CALM	229	44	396	138	807
CHOPPY	86	26	226	7	345
GLASS CALM	26	2	21		49
HIGH	2		8	4	14
MOD/CHOP	67	31	251	144	493
ROUGH	64	18	162	48	292
SLIGHT	261	89	648	211	1,209
SMOOTH	209	49	371	134	763
VERY HIGH		1			1
VERY ROUGH	10	2	6		18
(blank)		1			1
Grand Total	**954**	**263**	**2,089**	**686**	**3,992**

[3] https://data-rnli.opendata.arcgis.com/datasets/rnli-returns-of-service/explore
[4] https://data-rnli.opendata.arcgis.com/datasets/rnli-returns-of-service/explore

Attributes included in the RNLI dataset include:

- Latitude and Longitude of the rescue with some missing values
- AIC (Description of the Incident – over 200 different values). This is suitable for text mining in a future project, not used in this project
- Lifeboat station. This value was used to lookup the country and region of the lifeboat station. The country was used to determine which bank holidays applied. The station was also used to determine which weather station to use for weather data.
- RoS Type – What type of boat was used to respond to the incident
- CasualtyCategory, CasualtyTypeFull, Reason for Launch and Outcome of Service – These are categorisation field that were grouped for analysis.
- Visibility, weather and sea conditions – at Launch and at the Incident. These are dependent on crew observations and was replaced with Weather from Met Eireann.
- Date and Time of Lunch. The date was used to determine the season as there was seasonality in related data. The time was used to split the data between morning, afternoon, evening, and night.

3.2 Other Data

Weather data: Hourly Irish weather up to 2020 as consolidated on Kaggle[5].The hours were updated from 24:00:00 to 00:00:00 to reflect the standard SQL format.
Bank Holidays in Ireland[6] were used to determine bank holiday weekends.
Calendar was scripted in hourly intervals.

3.3 Mapping weather station data to RNLI stations

The Haversine algorithm calculates the distance between two objects on the surface of a sphere (Prasetya et al., 2020) and is the best option for distances between two co-ordinates.

Palupi, et al (2021) compared the results of the distance between two locations using the Haversine Formula and the Euclidean Formula. They found that over shorter distances the difference between the two formulas was not statistically significant. Another comparison of the Euclidean and Haversine methods found that the difference over a short distance is only 0.003% (Maria et al., 2020)
The Euclidean method was therefore used to determine the distance between the Irish Lifeboat Stations and the Irish Weather Stations

The difference was calculated as follows:
For each RNLI Station:

- Calculate the difference with each weather station
- Difference = $\sqrt{[(\text{Latitude RNLI Station} - \text{Latitude Weather Station})^2 + (\text{Longitude RNLI Station} - \text{Longitude Weather Station})^2]}$

[5] https://www.kaggle.com/conorrot/irish-weather-hourly-data
[6] https://www.citizensinformation.ie/en/employment/employment_rights_and_conditions/leave_and_holidays/public_holidays_in_ireland.html#

- The Weather station with the smallest value is the closest to the RNLI Station
- Review completeness of weather station data. Mount Dillon missed 3,312 (3.15%) of hourly values. Finner 1,301 (1.24%) of hourly values. Not material enough to impact model.

3.4 Combining data sets

- The 36 RNLI stations were combined with the scripted data and times: 36 stations x 5 years (365*4+366) x 24 hours per day = 1,577,664 records.
- The weather data for the closest station was added allowing the use of estimated, not actual, weather data as per related work (Zeng et al., 2020)
- The bank holidays were added
- The incidents were added.

Initial models had an accuracy in excess of 98% as less than 2% of hourly measure had an incident. The models simply predicted a "no" outcome for all parameters.

3.5 Grouping the data – Categorisation to improve the models

Based on the limitations and recommendations from the related work (Harikrishnan & Postwala, 2021; Pitman et al., 2019; Zou et al., 2020) the dataset was then reduce in complexity by grouping the data into categories.

Temperature:

- Very Cold <6
- Cold 6 – 15
- Moderate / Warm 16+

Windspeed[7] ,

- Light Winds - Wind speeds of 1 - 14 knots (1-16 mph or 1-26 km/h)
- Moderate Winds - Wind speeds of 15 - 19 knots (17-22 mph or 28-35 km/h)
- Strong Winds - Wind speeds of 20 - 33 knots (24-37 mph or 39-54 km/h)
- Gale - Wind speeds of 34 - 47 knots (39-54 mph or 63-87 km/h)
- Storm - Wind speeds of 48 - 63 knots (55-73 mph or 89-117 km/h)

Visibility[8]

- Good: more than 5 nautical miles (9km)
- Moderate: 2 – 5 nm (4 – 9 km)
- Poor: 0.5 to 2 nm (4km)
- Fog: less than 0.5 nm (1,000m)

[7] https://www.safeboater.com/learn-the-rules/weather.html
[8] https://www.met.ie/forecasts/marine-inland-lakes/sea-area-forecast-terminology

3.6 Modelling

By Using a tool like RapidMiner deployed on a Virtual Machine in the Azure Cloud multiple models can be analysed very quickly.
There was no consistency in the literature as to the best model to use, and a large number of varied models were reviewed with different models suggested as best fit in the related literature:Gradient Boosted Trees, Deep Learning, Support Vector Machine, Random Forest, Naive Bayes, Logistic Regression, Generalized Linear Model, Fast Large Margin, Decision Tree

3.7 Improving models

(Majeed, 2019) finds that models speed and performance improves greatly without a significant impact on accuracy by giving more weight to variables with a bigger impact. To that end the records were grouped by time of day – creating four categories per day instead of 24 hours.

- Morning
- Afternoon
- Evening
- Night

The models were then re-run.

4 Design Specification

4.1 Architecture

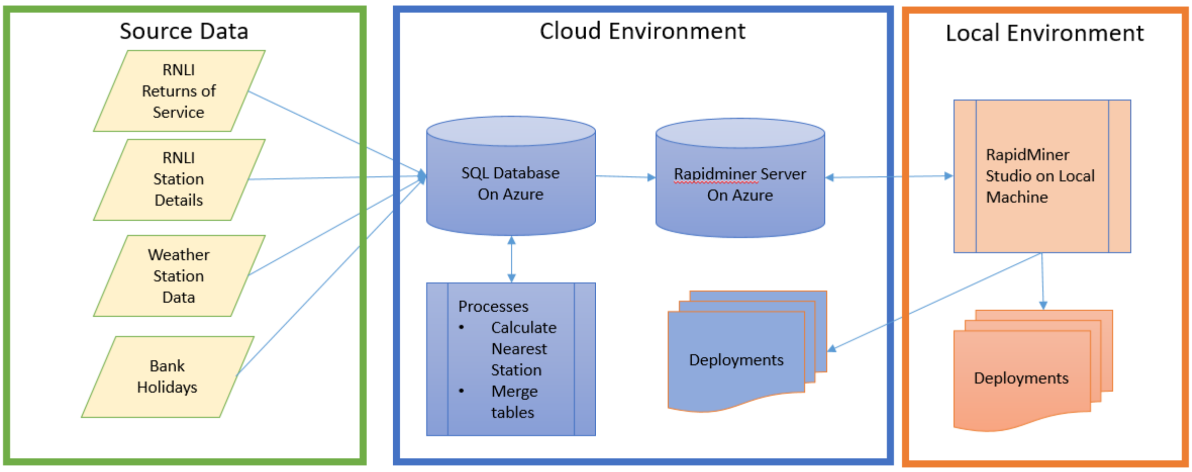

Figure 1: Project Architecture

ETL:

- Extraction: Data sources obtained were in the form of flat files (.csv).
- They were loaded into a SQL database via the import functionality
- The data was then transformed by merging tables and applying categorisation to the data

Initially this was done on a local machine, but it was later moved to the Cloud (Microsoft Azure) to allow for ease of use.

- The SQL Database was then connected to RapidMiner.

Processing Data

- Initial processing in RapidMiner was done by loading a flat file into a local repository.
- Subsequently a RapidMiner virtual machine was created in Azure and the data transferred to the RapidMiner AI Server for processing overnight.
- The interface is a local instance of RapidMiner Studio.
- The final models were deployed to the Cloud and for the purpose of demonstrating the model, also to my local machine.

4.2 Framework

The selected framework must include phases for discovery and analysis and include both processes and practices.(de Souza et al., 2019)

The CRoss Industry Standard Process for Data Mining (CRISP-DM) model was selected.

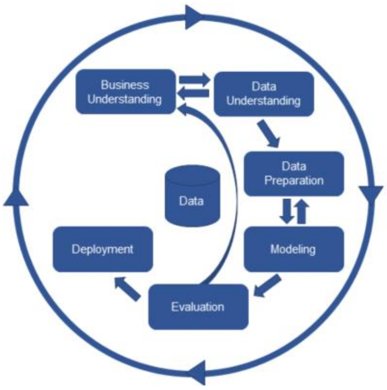

Figure 2: Phases of the CRISP-DM Process Model for Data Mining.

The benefit of the CRISP-DM Process model is that it is technology independent. (Wirth & Hipp, 2000) (Chapman et al., 2000).

Table 3: Application of the CRISP-DM Process Model

Phase	
Business Understanding	As a sailor and a member of the RNLI, this is an area of great interest to me. I also discussed the issues faced with active lifeboat volunteers in my local station. The ability to predict when possible incidents may occur is of great value to volunteers planning their daily activities.
Data Understanding	The Starting point was the datasets provided by the RNLI. As the purpose of this study is limited to the impact on volunteer time, i.e. launching boats, sensitive data related to the outcome of events and the demographic of people involved in events were excluded.

	Where possible data was grouped referring commonly used groupings such as Wind direction and speed[9] and Visibility[10] .
Data Preparation	The RNLI dataset was combined with a dataset containing hourly intervals. The 2020 data was excluded due to the inconsistency in the data caused by the impact of Covid were self-isolation measures impacted on both Commercial and Leisure activity (Lopes & Jaspal, 2020). Additional data was added in an iterative way as the models were refined. For example Bank Holidays.
Modeling	Various models were used to (1) categorise the data, and (2) predict whether an incident would occur
Evaluation	The models were evaluated based on accuracy.
Deployment	Deployment of the results in this project report. Future deployment could include providing access to volunteers after further refinement and evaluation.

4.3 Models and Algorithms

The models use include: Gradient Boosted Trees, Deep Learning, Support Vector Machine, Random Forest, Naive Bayes, Logistic Regression, Generalized Linear Model, Fast Large Margin, Decision Tree

Using RapidMiner for the modelling allows me to leverage well tested algorithms and focus on the data and evaluating the results.

5 Implementation

5.1 Data Transformation

The final transformed table contains the following attributes:

- BoatLaunched – yes or no – this is the value that we want to predict with our model
- Country – IRL – This is from the RNLI Station table.
- Region – Ireland or NI and Isle of Man
- HourGroupString – morning, afternoon, evening, or night
- Humidity – rounded to the nearest 10
- AvgTemp – the average of the Air, Wet Bulb and Dew Point temperatures, rounded to an integer
- Visibility – Good or Bad (less than 4km)
- Bank Holiday Weekend – yes or no – if there's a bank holiday on a Monday or Friday then the Saturday and Sunday are also considered as part of the Bank Holiday Weekend.

[9] https://www.safeboater.com/learn-the-rules/weather.html
[10] https://www.met.ie/forecasts/marine-inland-lakes/sea-area-forecast-terminology

- Wind S or W – yes or no – Winds that are Southerly or Westerly are associated with different weather to those from the North or East[11]
- Seasons [12] - summer, autumn, winter or spring. These are added as related work (Pitman et al., 2019) has found a seasonal impact
- Category, RoS Type, Outcome – these are grouped where there was an incident. Categorisation is subjective and can be improved by text analysis.

Table 4a: First 5 lines from the final data table – first 9 columns

Hour Group String	Humi-dity	Visi-bility Good v Bad	Avg Temp	Coun-try	Region	Bank Holiday Week-end	Season	Grouped Category
afternoon	20	Bad	8	IRL	Ireland	no	Spring	no incidents
afternoon	20	Bad	9	IRL	Ireland	no	Spring	no incidents
afternoon	30	Bad	6	IRL	Ireland	no	Spring	no incidents
afternoon	30	Bad	7	IRL	Ireland	no	Spring	no incidents
afternoon	30	Bad	8	IRL	Ireland	no	Spring	no incidents

Table 4b: First 5 lines from the final data table – last 6 columns

RoSType Grouped	Grouped Outcome	Boat Launched	WindSorW	Windspeed	Grouped Type
no incidents	no incidents	no	no	LightModerate	no incidents
no incidents	no incidents	no	no	LightModerate	no incidents
no incidents	no incidents	no	no	LightModerate	no incidents
no incidents	no incidents	no	no	LightModerate	no incidents
no incidents	no incidents	no	no	LightModerate	no incidents

5.2 Tools and Languages

RapidMiner is a language independent machine learning tool that can use data from multiple sources including SQL databases and flat files[13]. Recently RapidMiner data mining software has been used successfully in many research papers to predict market trends for precious metal prices, measure the performance of environmentally friendly coolants, classify breast cancer tumours, Evaluating User Satisfaction , Transformer Fault Analysis, etc using techniques such as Artificial Neural Networks (ANN), Decision Trees, SVM, k-NN, and Naïve Bayes (Çelik & Başarır, 2021),(Saengsikhiao & Taweekun, 2021), (Nanda & Jatain, 2021), (Uska et al., 2020) (Zhong et al., 2020)

[11] https://www.met.ie/climate/what-we-measure/wind
[12] https://www.nationalgeographic.org/encyclopedia/season/
[13]https://rapidminer.com/why-rapidminer/

A comparison of RapidMiner and using R found that new methods of analysis were available in R before RapidMiner and that the visualisation in R was better. However, RapidMiner has the benefit of being language independent, includes all established algorithms and the graphical user interface is very easy to use. (Dwivedi et al., 2016)
Kitcharoen, et al. (2013) found that RapidMiner worked well with large volumes of data in a cloud environment when they analysed production lines in the food industry.

(László & Ghous, 2020) Compared the results produced by Python and RapidMiner on two datasets and found similarities in the accuracy of the results.

5.3 Models Used

The following models were calculated as they were all referenced in the related works:

- Naïve Bayes
- Generalized Linear Model
- Logistic Regression
- Fast Large Margin
- Deep Learning
- Decision Tree
- Random Forest
- Gradient Boosted Trees
- Support Vector Machine

All models were automatically optimized and Columns were analysed in terms of Correlation and the importance of each column for each of the models.

Models were calculated several times with adjustments to data included, data categorisation, and model settings between calculations.

6 Evaluation

The analysis in RapidMiner produced detailed results for all the models. Figures 3a, 3b, and 3c below show the models results in terms of Accuracy, Precision and Cost. Taking only these graphs into account it is clear that the models are very similar in terms of accuracy, but there are large variances in precision and cost.

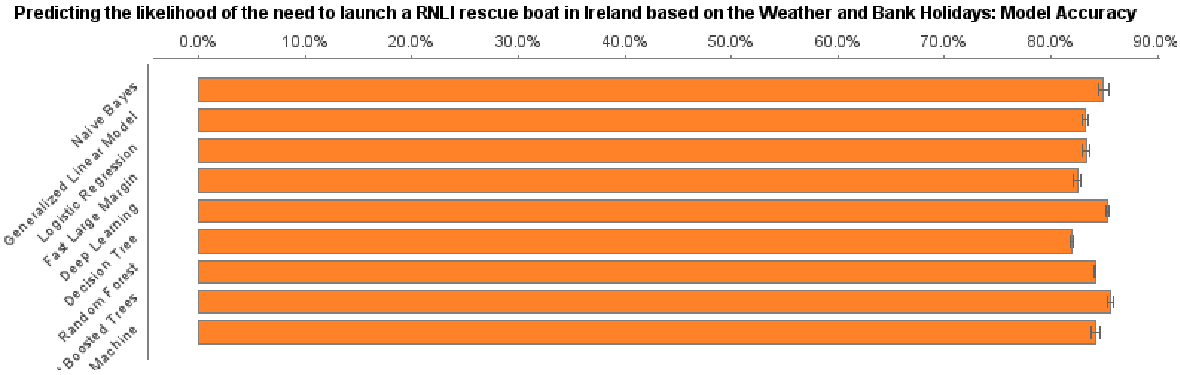

Figure 3: Results from RapidMiner: Model Accuracy

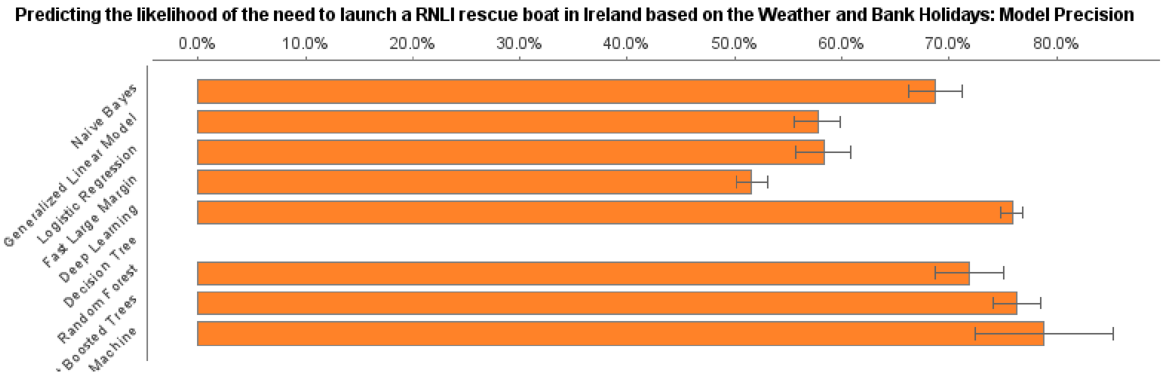

Figure 4: Results from RapidMiner: Model Precision

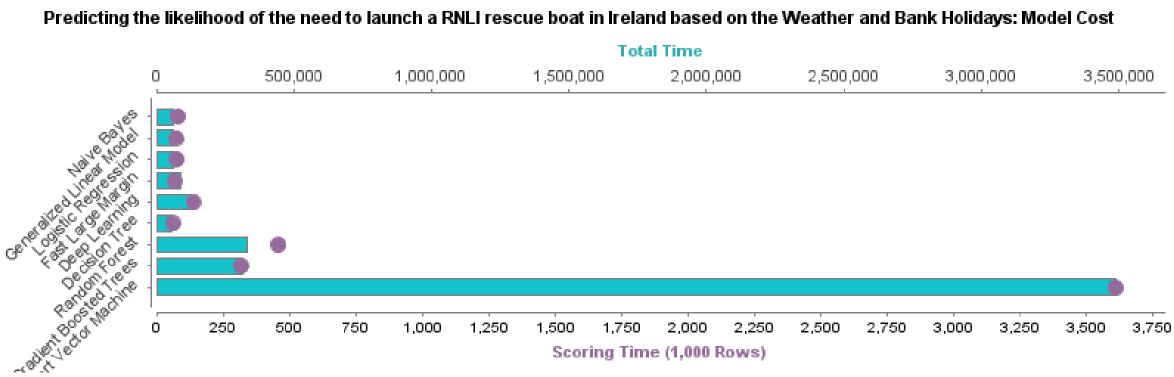

Figure 5: Results from RapidMiner: Processing cost

(Olson & Delen, 2008) recommends including the breakdown of accuracy by True Positives and True Negatives if, for example, a positive is more important. In this case it's important to be in the area in case of a launch and therefore the Class Precision from the Confusion Matrix is included in the summary of the results.

Table 5: Evaluating the Results from the various models

	Gradient Boosted Trees	Deep Learning	Support Vector Machine	Random Forest	Naive Bayes	Logistic Regression	Generalized Linear Model	Fast Large Margin	Decision Tree
Performance									
Accuracy %	85.6	85.3	84.5	84.1	85.0	83.3	83.2	82.5	82.0
Classification Error %	14.4	14.7	15.8	15.9	15.0	16.7	16.8	17.5	18.0
AUC %	85.9	85.5	83.0	83.5	83.9	81.9	81.9	82.7	50.0
Precision %	76.0	75.9	78.9	71.9	68.7	58.3	57.7	51.6	
Recall %	28.6	26.9	16.7	19.2	30.6	25.1	25.1	40.5	0.0
F Measure %	41.6	39.7	27.6	30.3	42.3	35.1	35.0	45.4	
Sensitivity %	28.6	26.9	16.7	19.2	30.6	25.1	25.1	40.5	0.0
Specificity %	98.0	98.1	99.0	98.3	96.9	96.1	96.0	91.7	100
Confusion Matrix									
Class Precision: No %	86.2	86.0	84.4	84.8	86.4	85.4	85.4	87.6	82.0
Class Precision: Yes %	76.3	75.9	78.7	71.7	68.7	58.4	57.8	51.7	0.0
Processing Cost									
Total Time (s)	312	128	3488	332	61	62	62	86	54

Accuracy & Classification Error: All models have an Accuracy over 82% which is considered good. The Classification error is the opposite of the Accuracy and therefore under 18% for all models, also consistently good.
AUC: The Decision Tree model has an AUC of 50% which indicates that there is no discrimination in the model, and it cannot be relied on. This is confirmed by the 0% class precision for Yes values which indicates that the model predicts “no” for all values.
F Measure: this measure should be viewed in conjunction with other measures and in this case it’s low for Random Forest and SVM.
Class Precision for a “yes” value, i.e. days on which the boats are launched are predicted correctly. This is only over 70% for Deep Learning, Gradient Boosted Trees, Random Forest and Support Vector Machine.
Processing Time: As the model is intended for a deployment in Cloud the cost in terms of time is also of importance. The support vector machine is not a good option as it took 58 minutes to process. This is not suitable in a Cloud environment (Harikrishnan & Postwala, 2021; Viellechner & Spinler, 2020)

Based on a combination accuracy, class precision, F measure, and processing cost the two models with the best fit are the Gradient Boosted Trees and Deep Learning Models

Table 6: Weight and Ranking of Attributes

	Gradient Boosted Trees		**Deep Learning**	
Attribute	**Weight**	**Ranking**	**Weight**	**Ranking**
Windspeed	0.16	1	0.15	2
HourGroupString	0.12	2	0.11	3
Humidity	0.11	3	0.16	1
BankHolidayWeekend	0.11	4	0.10	4
VisibilityGoodvBad	0.11	5	0.10	5
AvgTemp	0.07	6	0.07	6
Season	0.04	7	0.04	7
WindSorW	0.02	8	0.04	8

6.1 Gradient Boosted Trees

Figure 6 below shows the production model of the Gradient Boosted Trees as calculated.

On its own a decision tree cannot predict the weather or not a boat will have to be launched with any precision. When we combine several trees in a Gradient Boosted Tree model, the model calculates the improvement in the overall model by adding more and more trees. This is shown in the Gains/Lift table in the below figure.

We can see that the performance settings included a smaller decision tree (in terms of depth) of only 4 levels. A smaller tree is more generalizable and less likely to suffer from overfit[14]. It is also easier to explain to users. By combining 150 such trees the gradient boosted trees model can predict a launch (yes outcome) with a 76% precision.

Rather than interpreting individual trees we look at the overall importance of each variable to the ensemble of trees. This tells us which variables influence the outcome of the model more, and therefore we can determine which items impact the likelihood of having to launch a boat the most.

Attributes and ranking

Windspeed, the time of day and the humidity measure are biggest impact on the model, followed by the Bank holiday weekend.

It shows that there is a relationship between the weather (especially the windspeed and humidity) and the likelihood of an incident occurring that requires the launch of a boat.

[14] Class Notes from Postgraduate Diploma in Data Analytics at NCIrl during 2019. Moodle site no longer available.

In addition, we can see that there's a relationship between bank holidays and the number of incidents. If we consider Bank holiday as a proxy for the number of people out and about then the results are consistent with previous studies that found a relationship between the number of people on the beach and incidents of drowning.

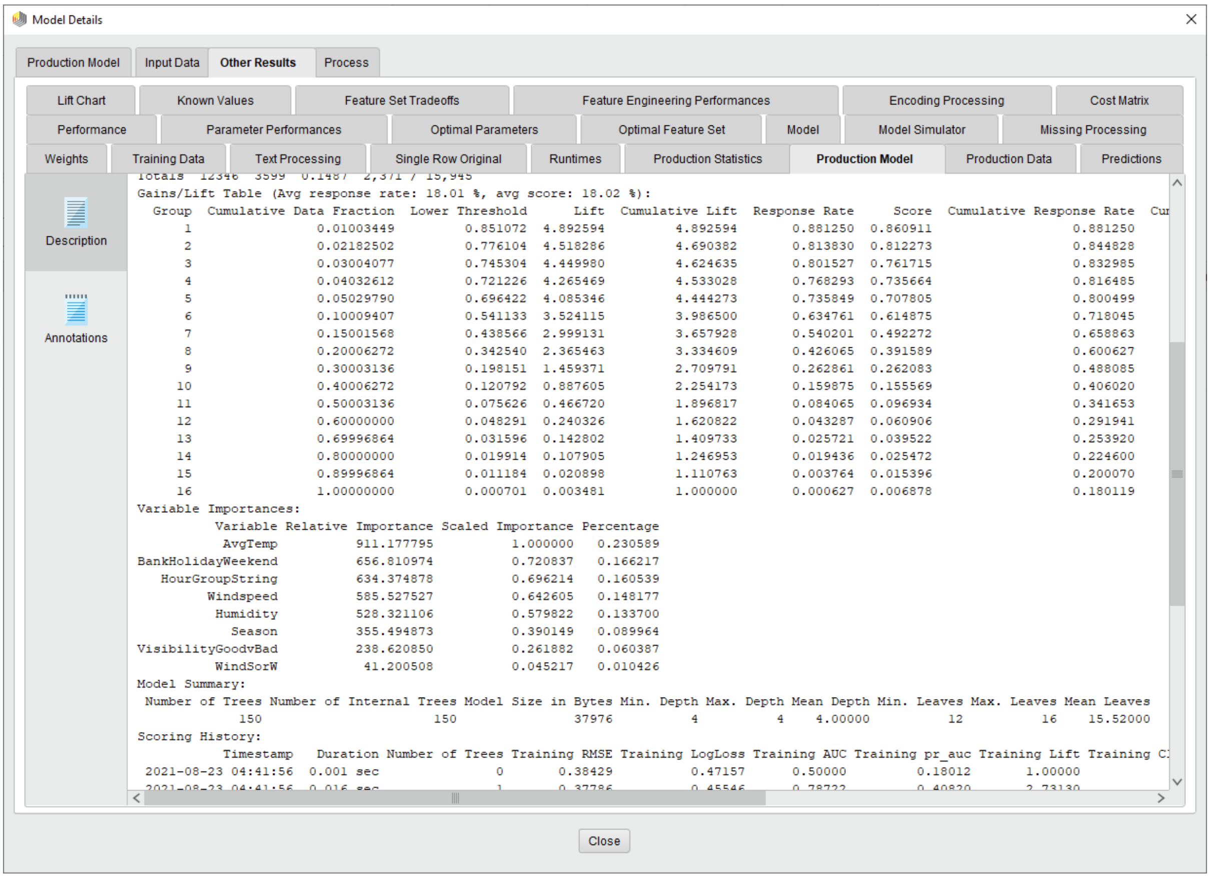

Figure 6: Gradient Boosted Tree Model.

6.2 Deep Learning

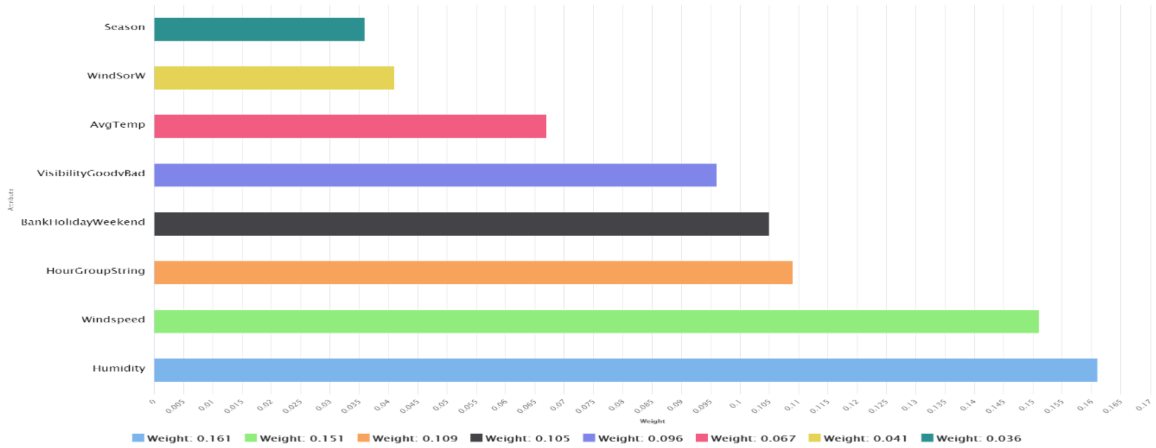

Figure 7: Deep Learning Model – weight of variables in the production model

Figure 7 shows the relative weight of the variables described in Table 6 in a visual graph. This is part of the production model generated by RapidMiner and displayed in figure 8 below. Humidity has a much higher weight of 16% in this model compared to the 11% in the Gradient Trees Model.

The deep learning model in a complex artificial neural network of four layers as can be seen in figure 8 below. The results from the model include the values used for evaluation and the Gains/Lift table showing how the model was built. As neural networks have hidden layers there's no simple algorithm to display. When deployed the predictions are based on the full network.

Deep Learning Model

```
Model Metrics Type: Binomial
 Description: Metrics reported on full training frame
 model id: rm-h2o-model-model-316
 frame id: rm-h2o-frame-model-316
 MSE: 0.09654792
 RMSE: 0.3107216
 R^2: 0.34640247
 AUC: 0.87932485
 pr_auc: 0.6340769
 logloss: 0.3133979
 mean_per_class_error: 0.204041
 default threshold: 0.2848600149154663
 CM: Confusion Matrix (Row labels: Actual class; Column labels: Predicted class):
         no   yes   Error          Rate
    no 6786  1058  0.1349  1,058 / 7,844
   yes  471  1253  0.2732    471 / 1,724
Totals 7257  2311  0.1598  1,529 / 9,568
Gains/Lift Table (Avg response rate: 18.02 %, avg score: 17.56 %):
  Group  Cumulative Data Fraction  Lower Threshold      Lift  Cumulative Lift  Response Rate     Score  Cumulative Response Rate  Cumulative Score  Capture Rate  Cumulative
      1                0.01013796         0.813199  4.863300         4.863300       0.876289  0.843809                  0.876289          0.843809      0.049304
      2                0.02006689         0.775246  4.439907         4.653809       0.800000  0.793354                  0.838542          0.818844      0.044084
      3                0.03010033         0.741038  4.162413         4.490010       0.750000  0.758081                  0.809028          0.798590      0.041763
      4                0.04002926         0.717923  4.323068         4.448602       0.778947  0.730066                  0.801567          0.781593      0.042923
      5                0.05006271         0.689142  3.873357         4.333312       0.697917  0.703360                  0.780793          0.765914      0.038863
      6                0.10033445         0.564998  3.611463         3.971636       0.650728  0.625402                  0.715625          0.695512      0.181555
      7                0.15008361         0.439322  2.903196         3.617473       0.523109  0.503149                  0.651811          0.631748      0.144432
      8                0.20004181         0.351060  2.089914         3.235983       0.376569  0.392298                  0.583072          0.571948      0.104408
      9                0.30006271         0.200727  1.484608         2.652191       0.267503  0.272154                  0.477882          0.472017      0.148492
     10                0.39997910         0.110041  0.777913         2.183989       0.140167  0.151865                  0.393520          0.392041      0.077726
     11                0.50000000         0.064572  0.574126         1.861949       0.103448  0.085592                  0.335493          0.330739      0.057425
     12                0.60002090         0.036693  0.260966         1.595072       0.047022  0.049693                  0.287406          0.283890      0.026102
     13                0.69993729         0.020618  0.203186         1.396380       0.036611  0.027985                  0.251605          0.247359      0.020302
     14                0.79995819         0.010573  0.127584         1.237739       0.022989  0.015316                  0.223021          0.218346      0.012761
     15                0.89997910         0.004400  0.081190         1.109203       0.014629  0.007206                  0.199861          0.194881      0.008121
     16                1.00000000         0.000005  0.017398         1.000000       0.003135  0.001999                  0.180184          0.175588      0.001740
Status of Neuron Layers (predicting BoatLaunched, 2-class classification, bernoulli distribution, CrossEntropy loss, 4,202 weights/biases, 54.5 KB, 95,680 training samples,
 Layer Units      Type Dropout       L1       L2 Mean Rate Rate RMS Momentum Mean Weight Weight RMS Mean Bias Bias RMS
     1    30     Input  0.00 %
     2    50 Rectifier       0 0.000010 0.000000  0.404594 0.482461 0.000000   -0.021024   0.180720  0.283591 0.085822
     3    50 Rectifier       0 0.000010 0.000000  0.092476 0.146949 0.000000   -0.035039   0.150663  0.694246 0.211640
     4     2   Softmax         0.000010 0.000000  0.004262 0.003421 0.000000    0.042426   0.395457  0.000003 0.083752
Scoring History:
           Timestamp   Duration Training Speed    Epochs Iterations      Samples Training RMSE Training LogLoss Training r2 Training AUC Training pr_auc Training Lift Trainin
 2021-08-23 04:28:11  0.000 sec                  0.00000          0     0.000000           NaN              NaN         NaN          NaN             NaN           NaN
```

Figure 8: Deep Learning Model – extract from model metrics

Overall, the same 3 attributes have the highest weight, and the bottom 5 attributes are in the same order. This confirms the relationship exists between the weather, Bank Holidays, and the number of incidents.

6.3 Discussion

The data used has limitations in terms of scope. Only data from 2015 to 2019 was used. Additional years can be added to further improve the models. There is also only data from the RNLI. The model would greatly benefit from adding in details from the other volunteer rescues in the country.

It would also improve the model to add in additional measures of how busy the beaches would be – perhaps school holidays combined with travel data from the tourism board.

Further developing the model by adding the type of launch and the type of incident will be useful in terms of prevention. For example, if similar results to the UK (Pitman et al., 2019) were found in relation to the number of incidents related to onshore fishing it could inform an Irish Water Safety campaign to encourage people fishing onshore to wear lifejackets or buoyancy aids.

7 Conclusion and Future Work

The objective of this project is to describe the weather conditions prevalent during a launch and ultimately predict the likelihood of a launch given the weather. The following research questions were addressed:

- Can we predict the likelihood of the need to launch an Irish rescue boat based on the weather?
- Can we improve the prediction by adding in additional variables to estimate number of people?

We can predict the likelihood of the need to launch by creating a model with both the weather and an estimation of activity using Bank holiday data as a proxy.

Models were evaluated in terms of Accuracy, Class Prediction and Cost to Compute as any such model will require a Cloud deployment to be useful.
The two models with the best fit were the Gradient Boosted Trees model and the Deep Learning model.

There are limitations to this project in terms of scope. Future projects could expand the scope to include additional data and address these gaps. The geographic region could also be expanded to include Northern Ireland, and later the rest of the UK

(Kervick et al., 2016) found that the highest percentage of drownings were men. This may relate to the gender of water users, or a lack of care in a certain demographic. Adding demographic details to the analysis may inform future policy.

As the focus was on volunteer activity (i.e., launching the boats) the reason for, and outcome of, the launch was ignore. Adding this level of detail into future studies will inform water user education, reduce the number of launches, and ultimately save lives.

8 References

Bailey, S., Olivera-Villarroel, S. M., & Lyubchich, V. (2020). Impacts of Inclement Weather on Traffic Accidents in Mexico City. In V. Lyubchich, Y. R. Gel, K. H. Kilbourne, T. J. Miller, N. K. Newlands, & A. B. Smith (Eds.), *Evaluating Climate Change Impacts* (1 ed., pp. 18). Chapman and Hall/CRC.

Bibby, C., & Reid, I. (2005, 18-22 April 2005). Visual Tracking at Sea. Proceedings of the 2005 IEEE International Conference on Robotics and Automation,

Burnett, R. A., & Dong Si. (2017). *Prediction of Injuries and Fatalities in Aviation Accidents through Machine Learning* Proceedings of the International Conference on Compute and Data Analysis, Lakeland, FL, USA. https://doi.org/10.1145/3093241.3093288

Capucci, A., Aschieri, D., Piepoli, M. F., Bardy, G. H., Iconomu, E., & Arvedi, M. (2002). Tripling Survival From Sudden Cardiac Arrest Via Early Defibrillation Without Traditional Education in Cardiopulmonary Resuscitation. *Circulation*, *106*(9), 1065-1070. https://doi.org/10.1161/01.CIR.0000028148.62305.69

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-Step Data Mining Guide*. Retrieved 21 April from https://s2.smu.edu/~mhd/8331f03/crisp.pdf

de Souza, A. A., Barcelos, T. S., Munoz, R., Villarroel, R., & Silva, L. A. (2019). Data Mining Framework to Analyze the Evolution of Computational Thinking Skills in Game Building Workshops. *IEEE Access*, *7*, 82848-82866. https://doi.org/10.1109/ACCESS.2019.2924343

De Vos, J. (2020). The effect of COVID-19 and subsequent social distancing on travel behavior. *Transportation Research Interdisciplinary Perspectives*, *5*, 100121. https://doi.org/https://doi.org/10.1016/j.trip.2020.100121

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., . . . Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553-597. https://doi.org/https://doi.org/10.1002/qj.828

Dwivedi, S., Kasliwal, P., & Soni, S. (2016, 18-19 March 2016). Comprehensive study of data analytics tools (RapidMiner, Weka, R tool, Knime). 2016 Symposium on Colossal Data Analysis and Networking (CDAN),

Eliopoulou, E., Papanikolaou, A., & Voulgarellis, M. (2016). Statistical analysis of ship accidents and review of safety level. *Safety Science*, *85*, 282-292. https://doi.org/https://doi.org/10.1016/j.ssci.2016.02.001

Faghih-Roohi, S., Xie, M., & Ng, K. M. (2014). Accident risk assessment in marine transportation via Markov modelling and Markov Chain Monte Carlo simulation. *Ocean Engineering*, *91*, 363-370. https://doi.org/https://doi.org/10.1016/j.oceaneng.2014.09.029

Gallagher, S., Tiron, R., Whelan, E., Gleeson, E., Dias, F., & McGrath, R. (2016). The nearshore wind and wave energy potential of Ireland: A high resolution assessment of availability and accessibility. *Renewable Energy*, *88*, 494-516. https://doi.org/https://doi.org/10.1016/j.renene.2015.11.010

Grey, C., & O'Toole, M. (2020). The Placing of Identity and the Identification of Place: "Place-Identity" in Community Lifeboating. *Journal of Management Inquiry*, *29*(2), 206-219. https://doi.org/10.1177/1056492618768696

Harikrishnan, R., & Postwala, B. C. (2021). Predicting Road Accident Severity Due to Weather Conditions Using Classification Algorithms. In T. K. Sharma, C. W. Ahn, O. P. Verma, & B. K. Panigrahi, *Soft Computing: Theories and Applications* Singapore.

Kervick, A., Dix, D., Walters, T., Sweeney, R., & Sarma, K. (2016). 832 Examining coastal drowning fatalities in the Republic of Ireland, 2010–2013. *Injury Prevention*, *22*(Suppl 2), A297. https://doi.org/10.1136/injuryprev-2016-042156.832
Kim, S., Lee, J., & Yoon, T. (2021). Road surface conditions forecasting in rainy weather using artificial neural networks. *Safety Science*, *140*, 105302. https://doi.org/https://doi.org/10.1016/j.ssci.2021.105302
Kitcharoen, N., Kamolsantisuk, S., Angsomboon, R., & Achalakul, T. (2013, 29-31 May 2013). RapidMiner framework for manufacturing data analysis on the cloud. The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE),
Knecht, W. R. (2013). The "killing zone" revisited: Serial nonlinearities predict general aviation accident rates from pilot total flight hours. *Accident Analysis & Prevention*, *60*, 50-56. https://doi.org/https://doi.org/10.1016/j.aap.2013.08.012
Li, Z., Liu, P., Wang, W., & Xu, C. (2012). Using support vector machine models for crash injury severity analysis. *Accident Analysis & Prevention*, *45*, 478-486. https://doi.org/https://doi.org/10.1016/j.aap.2011.08.016
Lopes, B. C. d. S., & Jaspal, R. (2020). Understanding the mental health burden of COVID-19 in the United Kingdom. *Psychological Trauma: Theory, Research, Practice, and Policy*, *12*(5), 465-467. https://doi.org/10.1037/tra0000632
László, K., & Ghous, H. (2020). Efficiency Comparison of Python and RapidMiner. *Multidisciplinary Sciences*, *10*(3), 212-220. https://doi.org/ https://doi.org/10.35925/j.multi.2020.3.26
Macleod, R. (2018). RNLI: 'Respect the Water' Campaign – Putting Prevention First. *International Journal of Aquatic Research and Education*, *11*(2, Article 11), Available at: https://scholarworks.bgsu.edu/ijare/vol11/iss12/11.
Majeed, A. (2019). Improving Time Complexity and Accuracy of the Machine Learning Algorithms Through Selection of Highly Weighted Top k Features from Complex Datasets. *Annals of Data Science*, *6*(4), 599-621. https://doi.org/10.1007/s40745-019-00217-4
Maria, E., Budiman, E., Haviluddin, & Taruk, M. (2020). Measure distance locating nearest public facilities using Haversine and Euclidean Methods. *Journal of Physics: Conference Series*, *1450*, 012080. https://doi.org/10.1088/1742-6596/1450/1/012080
Mohamed, E. A. (2014). Predicting Causes of Traffic Road Accidents Using Multi-class Support Vector Machines. *Journal of Communication and Computer*, *11*, 441-447. https://doi.org/10.17265/1548-7709/2014.05
Mondal, A. R., Bhuiyan, M. A. E., & Yang, F. (2020). Advancement of weather-related crash prediction model using nonparametric machine learning algorithms. *SN Applied Sciences*, *2*(8), 1372. https://doi.org/10.1007/s42452-020-03196-x
Morgan, D., & Ozanne-Smith, J. (2013). Surf Lifeguard Rescues. *WILDERNESS & ENVIRONMENTAL MEDICINE*, *24*, 285–290.
Mullai, A., & Paulsson, U. (2011). A grounded theory model for analysis of marine accidents. *Accident Analysis & Prevention*, *43*(4), 1590-1603. https://doi.org/https://doi.org/10.1016/j.aap.2011.03.022
Munro, R. (2011). Working on water [Article]. *Nursing Standard*, *25*(44), 20-22. https://doi.org/10.7748/ns.25.44.20.s25
Nanda, A., & Jatain, A. (2021, 2021//). Analysis of Breast Cancer Detection Techniques Using RapidMiner. Proceedings of International Conference on Artificial Intelligence and Applications, Singapore.

Ni, X., Wang, H., Che, C., Hong, J., & Sun, Z. (2019). Civil aviation safety evaluation based on deep belief network and principal component analysis. *Safety Science*, *112*, 90-95. https://doi.org/https://doi.org/10.1016/j.ssci.2018.10.012

Olson, D. L., & Delen, D. (2008). Performance evaluation for predictive modeling. In *Advanced Data Mining Techniques* (pp. 180). Springer.

Oluseye, O. O., & Ogunseye, O. O. (2016). Human Factors as Determinants of Maritime Accidents in Maritime Companies in Nigeria. *Journal of Maritime Research*, *XIII*(III), 61-68.

O'Toole, M., & Calvard, T. (2019). I've Got Your Back: Danger, Volunteering and Solidarity in Lifeboat Crews. *Work, Employment and Society*, *34*(1), 73-90. https://doi.org/10.1177/0950017019862962

O'Toole, M., & Grey, C. (2015). Beyond choice: 'Thick' volunteering and the case of the Royal National Lifeboat Institution. *Human Relations*, *69*(1), 85-109. https://doi.org/10.1177/0018726715580156

Palupi, R., Yulianna, D., & Winarsih, S. (2021). Analisa Perbandingan Rumus Haversine Dan Rumus Euclidean Berbasis Sistem Informasi Geografis Menggunakan Metode Independent Sample t-Test (Translated to English from Bahasa Indonesia). *JITU : Journal Informatic Technology And Communication, 5(1)*, *5*(1), 40-47. https://doi.org/10.36596/jitu.v5i1.494.

Pitman, S. J., Wright, M., & Hocken, R. (2019). An analysis of lifejacket wear, environmental factors, and casualty activity on marine accident fatality rates. *Safety Science*, *111*, 234-242. https://doi.org/https://doi.org/10.1016/j.ssci.2018.07.016

Prasetya, D. A., Nguyen, P. T., Faizullin, R., Iswanto, I., & Armay, E. F. (2020). Resolving the Shortest Path Problem using the Haversine Algorithm. *Journal of Critical Reviews*, *7*(1), 62-64.

Roger, K. (2004). Communities of Practice in the Royal National Lifeboat Institution. In H. Paul & K. Chris (Eds.), *Knowledge Networks: Innovation Through Communities of Practice* (pp. 70-78). IGI Global. https://doi.org/10.4018/978-1-59140-200-8.ch007

Royal National Lifeboat Institution. (2021). RNLI Annual Report and Accounts 2020. In (Vol. 2021, pp. 84). Poole, UK: RNLI.

Saengsikhiao, P., & Taweekun, J. (2021). The Data Mining Technique Using RapidMiner Software for New Zeotropic Refrigerant. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, *83*(1), 70-90. https://doi.org/10.37934/arfmts.83.1.7090

Seesink, J., Nieuwenburg, S. A. V., van der Linden, T., & Bierens, J. J. L. M. (2019). Circumstances, outcome and quality of cardiopulmonary resuscitation by lifeboat crews. *Resuscitation*, *142*, 104-110. https://doi.org/https://doi.org/10.1016/j.resuscitation.2019.07.012

Shih, G., Tsai, P., & Lin, C. (2018, 7-10 Oct. 2018). A Speed up Approach for Search and Rescue. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC),

Stokes, C., Masselink, G., Revie, M., Scott, T., Purves, D., & Walters, T. (2017). Application of multiple linear regression and Bayesian belief network approaches to model life risk to beach users in the UK. *Ocean & Coastal Management*, *139*, 12-23. https://doi.org/https://doi.org/10.1016/j.ocecoaman.2017.01.025

Tipton, M., Reilly, T., Rees, A., Spray, G., & Golden, F. (2008). Swimming Performance in Surf: The Influence of Experience. *Int J Sports Med.*, *29*(11), 895-898.

Tipton, M. J., & Golden, F. S. C. (2011). A proposed decision-making guide for the search, rescue and resuscitation of submersion (head under) victims based on expert opinion.

Resuscitation, *82*(7), 819-824. https://doi.org/https://doi.org/10.1016/j.resuscitation.2011.02.021

Toivonen, A. (2016). 830 Drowning from the viewpoint of the near-drowned. *Injury Prevention*, *22*(Suppl 2), A296. https://doi.org/10.1136/injuryprev-2016-042156.830

Uska, M., Wirasasmita, R., Usuluddin, U., & Arianti, B. (2020). Evaluation of Rapidminer-Aplication in Data Mining Learning using PeRSIVA Model. *Edumatic: Jurnal Pendidikan Informatika*, *4*, 164-171. https://doi.org/10.29408/edumatic.v4i2.2688

Viellechner, A., & Spinler, S. (2020, 2020-01-07). Novel Data Analytics Meets Conventional Container Shipping: Predicting Delays by Comparing Various Machine Learning Algorithms. Proceedings of the 53rd Hawaii International Conference on System Sciences, Hawaii, USA.

Ward, J., & Greene, A.-m. (2018). Too Much of a Good Thing? The Emotional Challenges of Managing Affectively Committed Volunteers. *Nonprofit and Voluntary Sector Quarterly*, *47*(6), 1155-1177. https://doi.org/10.1177/0899764018783276

Water Safety Ireland. (2021). *Community Rescue Boats Ireland*. Government of Ireland. Retrieved 8/8/2021 from https://watersafety.ie/community-rescue-boats-ireland/

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.

Xinqiang, C., Chaojian, S. H. I., Shengzheng, W., Huafeng, W. U., Tie, X. U., & Ke, R. (2017). Research Review on Marine Search and Rescue. *Journal of Shipping and Ocean Engineering*, *7*. https://doi.org/10.17265/2159-5879/2017.03.004

Zeng, Q., Hao, W., Lee, J., & Chen, F. (2020). Investigating the Impacts of Real-Time Weather Conditions on Freeway Crash Severity: A Bayesian Spatial Analysis. *International Journal of Environmental Research and Public Health*, *17*(8), 2768.

Zhong, J.-y., Wei, J.-x., Wu, X.-r., & Tang, H. (2020). Transformer Fault Diagnosis Based on RapidMiner and Modified ELM Algorithm. *Journal of Physics: Conference Series*, *1585*, 012030. https://doi.org/10.1088/1742-6596/1585/1/012030

Zou, Y., Zhu, T., Xie, Y., Li, L., & Chen, Y. (2020). Examining the Impact of Adverse Weather on Travel Time Reliability of Urban Corridors in Shanghai. *Journal of Advanced Transportation*, *2020*, 8860277. https://doi.org/10.1155/2020/8860277

Çelik, U., & Başarır, C. (2021). The Prediction of Precious Metal Prices via Artificial Neural Network by Using RapidMiner [Articles]. *5*. https://doi.org/https://dergipark.org.tr/en/pub/alphanumeric/290381