A

Mini-Project Report

on

Illegal Fishing Prediction System

Submitted as partial fulfilment for the award of

BACHELOR OF TECHNOLOGY DEGREE

Session 2022-23

in

Computer Science & Engineering

Submitted By:

RIYA GUPTA Roll.No:85

SHASHANK SRIVASTAVA Roll.No:100

Under the guidance of:

Ms. Kavita Sharma

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS COLLEGE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

AFFILIATED TO

DR. A.P.J ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW, UP

GALGOTIAS COLLEGE OF ENGINEERING & TECHNOLOGY
GREATER NOIDA, UTTER PRADESH, INDIA- 201306.

CERTIFICATE

This is to certify that the mini-project report entitled "**ILLEGAL FISHING PREDICTION SYSTEM**" submitted by Ms. **RIYA GUPTA (Roll.No:2000970100085)** and Mr. **SHASHANK SRIVASTAVA (Roll.No:2000970100100).** to the Galgotias College of Engineering & Technology, Greater Noida, Utter Pradesh, affiliated to Dr. A.P.J. Abdul Kalam Technical University Lucknow, Uttar Pradesh in partial fulfillment for the award of Degree of Bachelor of Technology in Computer science & Engineering is a bonafide record of the project work carried out by them under my supervision during the year 2022-2023.

Ms. Kavita Sharma	**Dr. Vishnu Sharma**
Assistant Professor	**Professor and Head**
Deptt. of CSE	**Deptt. of CSE**

GALGOTIAS COLLEGE OF ENGINEERING & TECHNOLOGY

GREATER NOIDA, UTTER PRADESH, INDIA- 201306.

ACKNOWLEDGEMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend my sincere thanks to all of them.

We are highly indebted to **Ms. Kavita Sharma** for her guidance and constant supervision. Also, we are highly thankful to him for providing necessary information regarding the project & also for his support in completing the project.

We are extremely indebted to **Dr. Vishnu Sharma**, HOD, Department of Computer Science and Engineering, GCET and **Ms. Kavita Sharma**, Project Coordinator, Department of Computer Science and Engineering, GCET for their valuable suggestions and constant support throughout my project tenure. We would also like to express our sincere thanks to all faculty and staff members of Department of Computer Science and Engineering, GCET for their support in completing this project on time.

We also express gratitude towards our parents for their kind co-operation and encouragement which helped me in completion of this project. Our thanks and appreciations also go to our friends in developing the project and all the people who have willingly helped me out with their abilities.

Riya Gupta

Shashank Srivastava

ABSTRACT

Fishing businesses facing a major financial problem throughout the globe due to illegal fishing. Through this illegal fishing, we are pushing many fish populations to extinction. In existed system they used data manipulation in illegal fishing data because of that there was a delay in catching the illegal vessels and in that system the data given is manual. In our paper, we are proposing data analytics to find these vessels. We can gather the primary data from the Global Fishing Watch (GFW), we analyse this data and find the vessels whether they are used for illegal fishing or legal. Based on the sensors attached to the vessel we can find AIS location data, type of the vessel, and speed of the vessel. By our model, we can predict illegal fishing and can take necessary actions against the illegal fishing boats.

Illegal fishing is prevalent throughout the world and heavily impacts the health of our oceans, the sustainability and profitability of fisheries, and even acts to destabilize geopolitical relations. To achieve the United Nations Sustainable Development Goal of "Life Below Water", our ability to detect and predict illegal fishing must improve. Recent advances have been made through the use of vessel location data, however, most analyses to date focus on anomalous spatial behaviours of vessels one at a time. To improve predictions, we develop a method inspired by complex systems theory to monitor the anomalous multi-scale behaviour of whole fleets as they respond to nearby illegal activities. Specifically, we analyse changes in the multiscale geospatial organization of fishing fleets operating on the Patagonia Shelf, an important fishing region with chronic exposure to illegal fishing. We show that legally operating (and visible) vessels respond anomalously to nearby illegal activities (by vessels that are difficult to detect). Indeed, precursor behaviours are identified, suggesting a path towards pre-empting illegal activities. This approach offers a promising step towards a global system for detecting, predicting and deterring illegal activities at sea in near real-time. Doing so will be a big step forward to achieving sustainable life underwater.

KEYWORDS: REGRESISON, WRANGLING, ANALYSIS, CORRELATION, OLS.

CONTENTS

LIST OF FIGURES

NOMENCLATURE

f(x)- Trained machine learning model

e- Exponential

y- slope

f1 score- Harmonic mean

Recall – TruePositives / (TruePositives + FalseNegatives)

Precision – TruePositives / (TruePositives + FalsePositives)

BAC- score of predicton

ABBREVIATIONS

LR- logistic regression

OLS- Ordinary Least Squares regression

FIT- fitting

SVC- Support Vector Machine

IUU- illegal, unreported, and unregulated

CLF- Classifier

RF- Random Forest

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION

Illegal fishing is one of the most dangerous phenomena that can destroy and affect the sea environment in a bad way especially in the pacific and Indian ocean. Moreover, illegal fishing is an activity of the fishers who don't have licenses and they are practicing their jobs without following the laws that established by the governments and they are using tools that have been banned because of the bad effects that those tools and materials cause. The reason for this is that The decline of the fish wealth that caused by illegal fishing is affecting badly million of people in the world who are depending on the fish and sea creatures and benefiting from them to survive. Consequently this thing causes a lot of diseases between the fish that can transfer to the human and affect their health's such as Heterophyids disease that affect the human stomach, salmonella and candida fungus that causes severe infections in some parts of the body. This shows that illegal fishing is not effecting the fish only it is also affecting other sea creatures such as coral reefs. The fishers that are practicing illegal fishing use tools and materials such as dynamite that are banned by the government to make it easy for them to get more fish in a short time.

5.7 DESCRIPTION

illegal fishing refers to fishing activities conducted by foreign vessels without permission in waters under the jurisdiction of another state, or which contravene its fisheries law and regulations in some other manner. Illegal fishing adversely impacts the health of our oceans and will continue to challenge our attempts to achieve sustainable ocean use unless it is addresses. Illegal fishing, typically referred to a more inclusive term, IUU (Illegal, Unreported, or Unregulated) fishing, is a global problem. Legal fishing alone depletes the fish populations to unsustainable levels; illegal fishing pushes many fish populations to the brink of extinction. With such a high global demand, there are attractive incentives for fisherman to go outside the law and fish for out of season fish or fish more than they are allowed. Along with very little global regulation black market fish easily spread into the general stock making them almost impossible to track once they are off the boat. To put the problem in perspective Black market fishing is estimated to account for 11 million – 26 million tonnage of fish equal to 14 or 33 per cent respectively of the world's total legal catch (fish and other marine fauna) in 2011. In the same year, legal fishing accounted for 78.9 million tonnage of fish. With such a potentially large subset of the global catch coming from IUU fishing being able to track and prevent it is a top priority (World Ocean View). With the advent of big data and new tracking system systems becoming available Lockheed Martin sees an opportunity to start becoming proactive on IUU fishing. Previously to find an illegal fishing vessel the Coast Guard or another entity would have to physically patrol the space. With such limited resources and such a wide area to search these tactics have been largely ineffective. Several large companies including Google have formed the Global Fishing Watch (GFW) to utilize this new tracking data to start to solve this problem. With the increased global focus and availability of data a real dent can start to be made in IUU fishing .

Illegal, Unreported, and Unregulated fishing?

Lllegal fishing : refers to fishing activities conducted in contravention of applicable laws and regulations, including those laws and rules adopted at the regional and international level.

Unreported fishing : refers to fishing activities that are not reported or are misreported to relevant authorities in contravention of national laws and regulations or reporting procedures of a relevant regional fisheries management organization.

Unregulated fishing : occurs in areas or for fish stocks for which there are no applicable conservation or management measures and where such fishing activities are conducted in a manner inconsistent with State responsibilities for the conservation of living marine resources under international law.

IUU fishing includes:

- Fishing without a license or quota for certain species.
- Failing to report catches or making false reports.
- Keeping undersized fish or fish that are otherwise protected by regulations.
- Fishing in closed areas or during closed seasons, and using prohibited fishing gear.
- Conducting unauthorized transhipments (e.g., transfers of fish) to cargo vessels.

1.3 BACKGROUND INFORMTION

The data set for the machine learning prediction of Illegal, Unreported, and Unregulated fishing is taken by the Global Fishing Watch's technology. Global Fishing Watch's open-access technology products, datasets, and code accelerate research and innovation and support sustainable ocean management. This map is the first open-access online platform for visualization and analysis of human activity at sea. Global Fishing Watch is working across the globe to provide governments and authorities with actionable reports and t h ervy building to help strengthen fisheries monitoring and compliance. Our global team of experts produce analyses to inform monitoring, control and surveillance of fisheries in four key areas:

- Illegal, unreported and unregulated fishing
- Transshipment
- Port controls
- Marine protected areas
- Fishing vessels.

1.4PROBLEM STATEMENT

Detecting IUU fishing is often a law enforcement human resource intensive task limited by funding. Newly available sources of data, along with models currently being developed have the potential to make a major difference in detecting illegal fishing activity by reducing and/or refocusing human resources on more productive investigations.

1.5PROBLEM DESCRIPTION

Currently there are very few models or analytics that exist for detecting illegal fishing without the physical search and seizure by law enforcement human resources. Modeling both fishing behavior and the illegal fishing enterprise will expose the data necessary to model and predict potential illegal fishing activity to focus law enforcement human resources to physically searching areas with a higher probability of detecting illegal activity. Thankfully major companies like Google are taking an interest into the problem. By using shipping vessels on board trackers, a methodology has started to be derived for describing typical fishing behavior using data from legal fishers. Multiple institutions are collecting and refining this data as well as developing algorithms to detect illegal activities.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

Literature review is a way through which we can find new ideas, concept. There are a lot of literatures published before on the same task; some papers are taken into consideration from which idea of the project is taken. This chapter comprises of detailed study done by our team on the topics related to our project like internet of things, led street lights and smart street light. The brief information for the same are discussed in the chapter ahead.

2.2 ILLEGAL FISHING

Illegal fishing is a serious problem that threatens the sustainability of fisheries around the world. Policy makers and fishery managers often rely on the imposition of strict sanctions and relatively intensive monitoring and enforcement programs to increase the costs of illegal behavior and thus deter it. However, while this can be successful in fisheries with sufficient resources to support high levels of surveillance and effective systems for imposing penalties, many fisheries lack the resources and requisite governance to successfully deter illegal fishing. Other types of governance systems, such as customary marine tenure and co-management, rely more on mechanisms such as norms, trust, and the perceived legitimacy of regulations for compliance. More generally, the absence of such social and psychological factors that encourage compliance in any fishery can undermine the efficacy of an otherwise effective and well-designed fishery management system. Here we describe insights from behavioral science that may be helpful in augmenting and securing the effectiveness of conventional deterrence strategies as well as in developing alternative means of deterring illegal fishing in fisheries in which high levels of surveillance and enforcement are not feasible. We draw on the behavioral science literature to describe a process for designing interventions for changing specific illegal fishing behaviors. The process begins with stakeholder characterization to capture existing norms, beliefs, and modes of thinking about illegal fishing as well as descriptions of specific illegal fishing behaviors. Potential interventions that may disrupt the beliefs, norms, and thought modes that give rise to these behaviors, along with those that encourage desirable behaviors, can be developed by applying principles gleaned from the behavioral science literature. These potential interventions can then be tested in artefactual experiments, piloted with small groups of actual stakeholders and, finally, implemented at scale.

2.3 DATA WRANGLING

Data wrangling is the process of cleaning and unifying messy and complex data sets for easy access and analysis. The readers of this article can expect to have a decent level of understanding of all the underlying phenomena and processes of data wrangling. The article will cover broader topics such as what data wrangling is all about, its associated processes, its importance in the field of data science, and essential tools involved in data wrangling, among other things.

2.3.1 Data Wrangling Tools and Techniques

There are several tools that can be used to perform data wrangling. The user must be proficient in whichever tool they use, as most of the time (~70%-80%) goes into data wrangling, and a decent skill-set is required. Among the most common tools for data wrangling are as follows-

- **Python**

Data wrangling in python can be performed with a lot of ease if one knows some basic libraries of python. Thus, one of the most popular languages, Python, can be used as a data wrangling tool. There are several techniques through which data wrangling in python can be performed. For example:

– Using pandas, tabula to convert it into a structured format

– Using csvkit, plotly to understand the data

– Using NumPy or pandas or clean and enrich the data

2.4 LIBRARIES

The Python Standard Library contains the exact syntax, semantics, and tokens of Python. It contains built-in modules that provide access to basic system functionality like I/O and some other core modules.Python Standard Library plays a very important role. Without it, the programmers can’t have access to the functionalities of Python. But other than this, there are several other libraries in Python that make a programmer’s life easier

Matplotlib: This library is responsible for plotting numerical data. And that’s why it is used in data analysis. It is also an open-source library and plots high-defined figures like pie charts, histograms, scatterplots, graphs, etc.

Pandas: Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data. Pandas support operations like Sorting, Re-indexing, Iteration, Concatenation, Conversion of data, Visualizations, Aggregations, etc.

Numpy: The name “Numpy” stands for “Numerical Python”. It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data. It consists of in-built mathematical functions for easy computations. Even libraries like TensorFlow use Numpy internally to perform several operations on tensors. Array Interface is one of the key features of this library.

SciPy: The name “SciPy” stands for “Scientific Python”. It is an open-source library used for high-level scientific computations. This library is built over an extension of Numpy. It works with Numpy to handle complex computations. While Numpy allows sorting and indexing of array data, the numerical data code is stored in SciPy. It is also widely used by application developers and engineers.

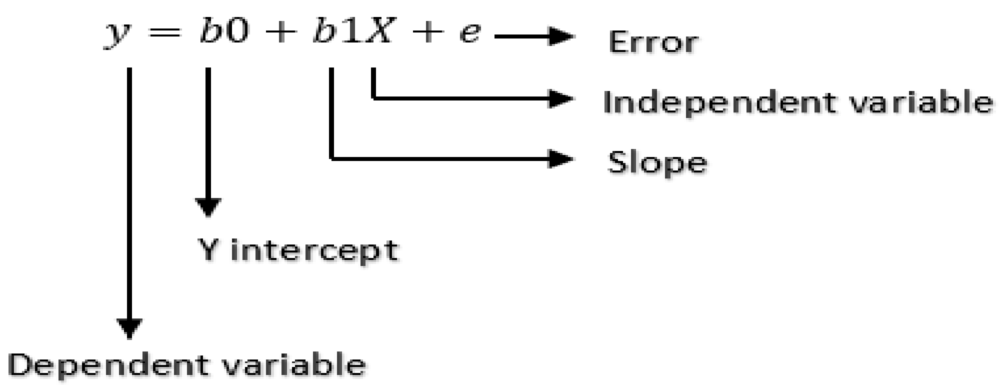
2.5 ALGORITHMS

2.5.1 REGRESSION:

Regression falls under the supervised learning category. The main goal of regression is the construction of an efficient model to predict the dependent attributes from a bunch of attribute variables. A regression problem is used when the output variable is either real or a continuous value i.e salary, score, weight, etc.

Linear Regression : Linear regression is one of the regression technique in which a dependent variable has a linear relationship with an independent variable. The main goal of Linear regression is to consider the given data points and plot the trend line that fit the data in the best way possible. Let’s say we have a dataset that contains information about the relationship between X and Y. Number of observations are made on X and Y and are recorded . This will be our training data. Our goal is to design a model that can predict the Y value if the X value is provided. Using the training data, a regression line is obtained which will give the minimum error. This linear equation is then used to apply for new data. That is, if we give X as an input, our model should be able to predict Y with minimum error.

The linear regression model is represented by the following equation:



After the model is built, We need to check the difference between the values predicted and actual data, if it is not much, then it is considered to be a good model. Below is a metric tool we can use to calculate errors in the model.

R — Square (R2) score:

$$R^2 = \frac{TSS - RSS}{TSS}$$

Where,

Total Sum of Squares (TSS): The measure of how a data set varies around a mean. The TSS tells us the variation in the dependent variable.

TSS = Σ (Y — Mean[Y])2

Residual Sum of Squares (RSS): sum of the squared differences between the actual Y and the predicted Y. The RSS tells us how much variation of the dependent variable is not explained by our model.

RSS = Σ (Y — f[Y])2

(TSS — RSS) measures the amount of variability in the response that is explained by performing the regression.

R2 score can be used to check all regression model's performance.

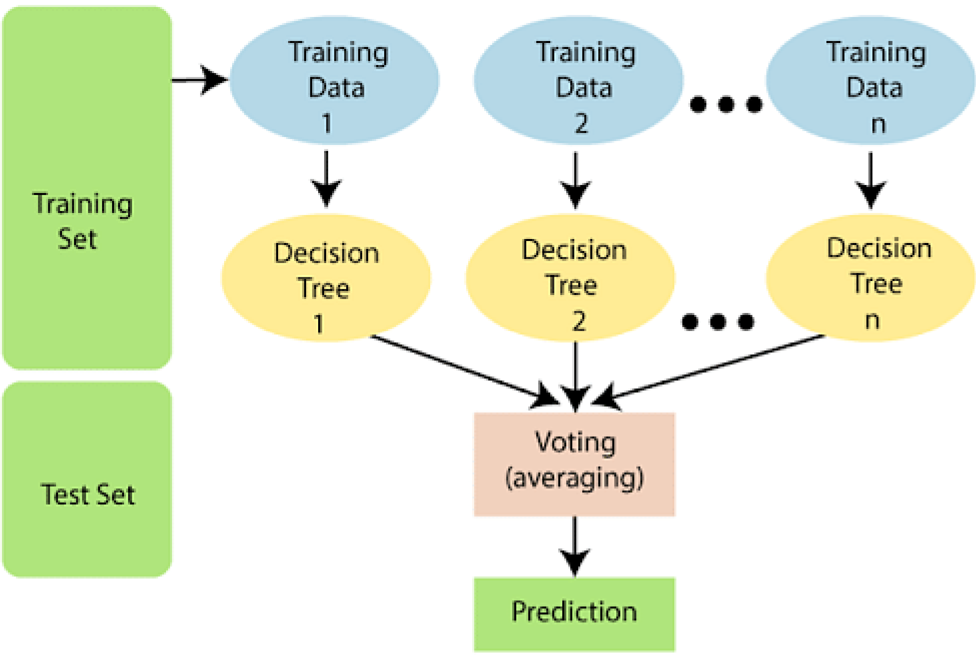
2.5.2 DECISION TREE:

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node. For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree. The complete process can be better understood using the below algorithm:

- **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.

- **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM).**
- **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
- **Step-4:** Generate the decision tree node, which contains the best attribute.
- **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

2.5.3 RANDOM FOREST



The following steps explain the working Random Forest Algorithm:

Step 1: Select random samples from a given data or training set.

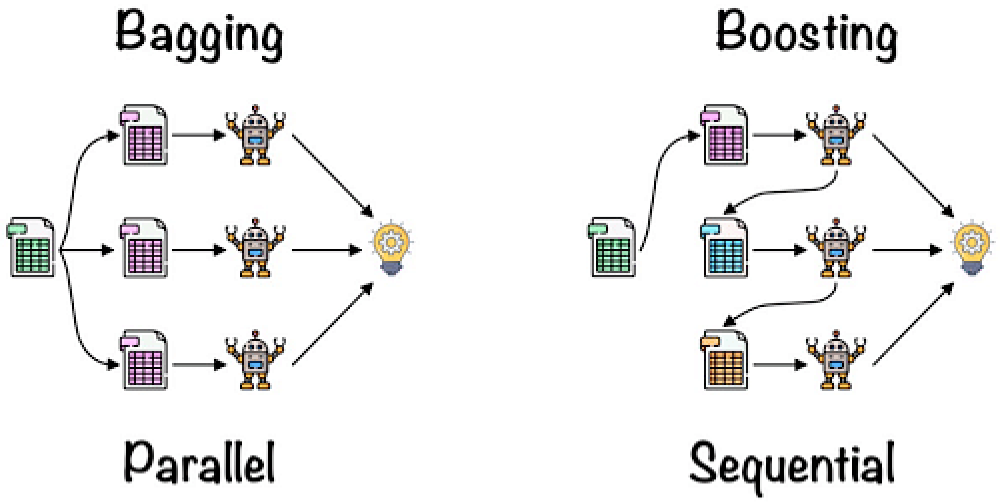
Step 2: This algorithm will construct a decision tree for every training data.

Step 3: Voting will take place by averaging the decision tree.

Step 4: Finally, select the most voted prediction result as the final prediction result.

This combination of multiple models is called Ensemble. Ensemble uses two methods:

1. **Bagging**: Creating a different training subset from sample training data with replacement is called Bagging. The final output is based on majority voting.
2. **Boosting**: Combing weak learners into strong learners by creating sequential models such that the final model has the highest accuracy is called Boosting. Example: ADA BOOST, XG BOOST.



CHAPTER 3: PROBLEM FORMULATION

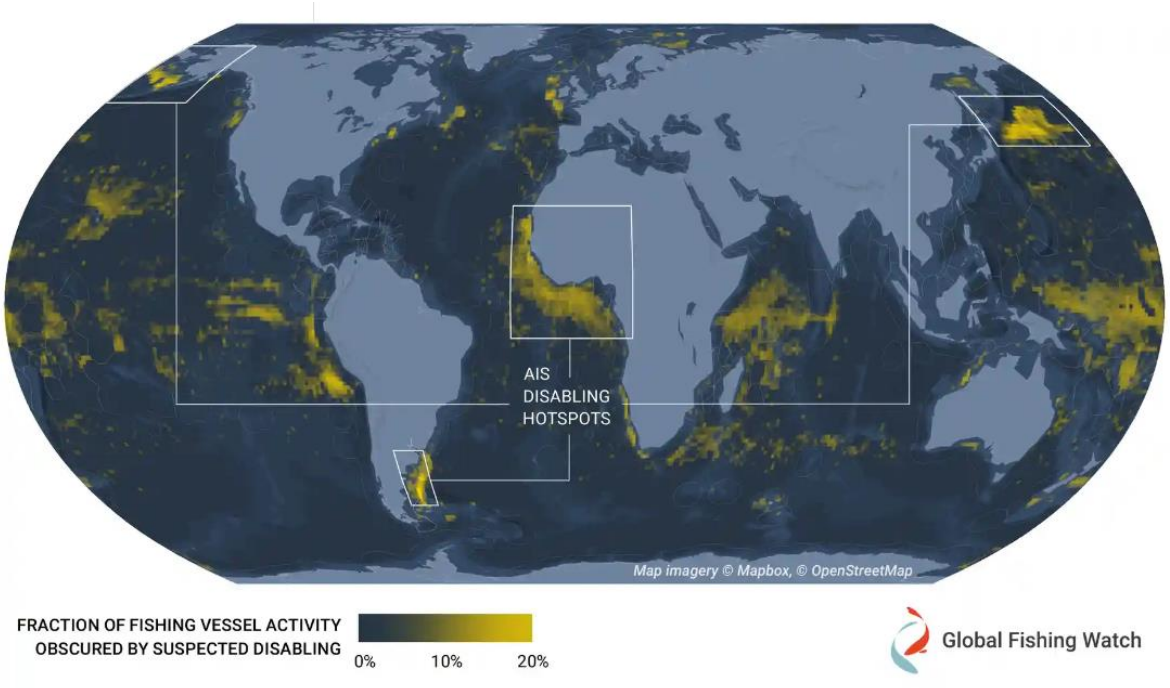
3.1 INRODUCTION

Problem formulation is the study and analysis of the problem for which the project was started. This chapter will be consisting of description of problem domain (which will give you an idea of why it is necessary to have illegal fishing prediction system), problem statement, the block diagram of our working model and the objectives of our project as if what is the aim of our project and how it is going to solve the list of problems listed in the problem statements.

3.2 DESCRIPTION OF PROBLEM DOMAIN

Illegal, unreported and unregulated (IUU) fishing is a broad term that captures a wide variety of fishing activity. IUU fishing is found in all types and dimensions of fisheries; it occurs both on the high seas and in areas within national jurisdiction. IUU fishing undermines national and regional efforts to conserve and manage fish stocks and, as a consequence, inhibits progress towards achieving the goals of long-term sustainability and responsibility. Moreover, IUU fishing greatly disadvantages and discriminates against those fishers that act responsibly, honestly and in accordance with the terms of their fishing authorizations. If IUU fishers target vulnerable stocks that are subject to strict management controls or moratoria, efforts to rebuild those stocks to healthy levels will not be achieved, threatening marine biodiversity, food security for communities who rely on fisheries resources for protein and the livelihoods of those involved in the sector Illegal, unreported, and unregulated (IUU) fishing significantly undermines the sustainability of the world's oceans.

According to the United Nations Food and Agriculture Organization estimates (FAO), 52% of the major marine fish stocks or species are fully exploited, 17% are overexploited, and 6% are depleted, and IUU fishing is one of the major contributors to this problem. The global scale of IUU fishing is estimated at about 11–26 million tons, which is about a $10–23.5 billion loss annually (MRAG, 2008). Some scientists suggest, that if current rates of depletion persist, most large predatory fish stocks will have collapsed by 2048 (Worm et al., 2006).



Illegal fishing vessels frequently use destructive fishing methods, and this harms crucial components of the marine ecosystem. Examples of these methods include blast bombing and cyanide fishing. Blast bombing has led to the loss of over 50% of the coral reef system in South East Asia (Caldwell and Fox, 2006), and has reduced the productive capacity of coral reefs to one fifth of their original capacity (White et al., 2000). The poisonous substance of cyanide fishing kills coral polyps, and the damage of these polyps leads to the discoloration of coral colonies (Mak et al., 2005).

3.3 PROBLEM FACED

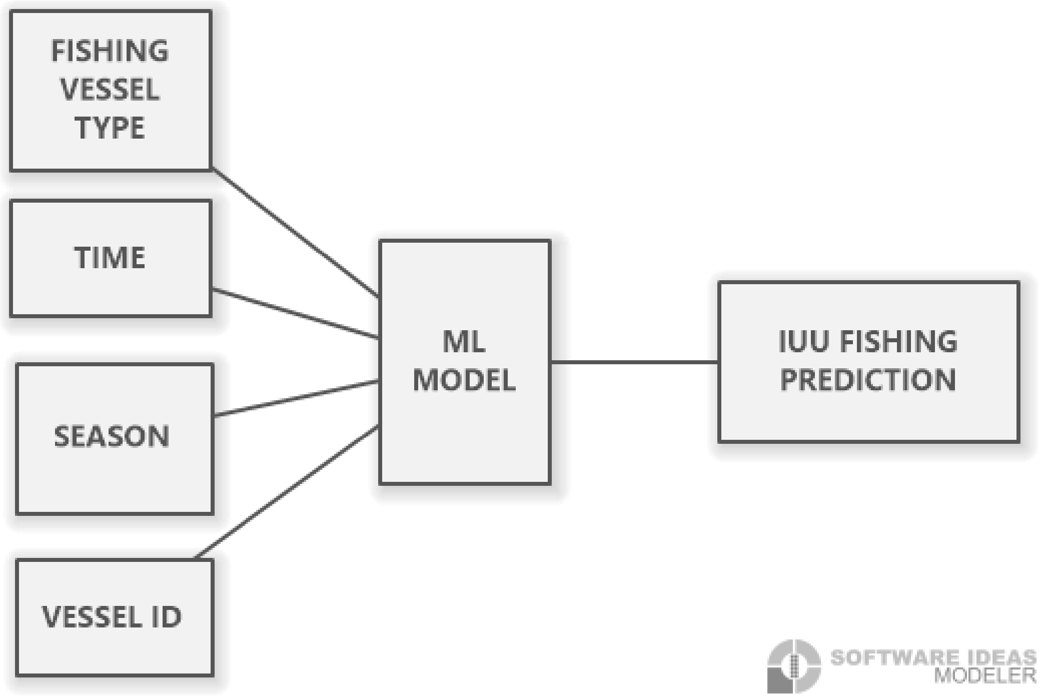
According to the Food and Agriculture Organisation (FAO), there are an estimated 4.6 million fishing vessels. Each vessel typically reports on at least 120 types of information, and during its lifetime one vessel may accumulate several information entries each time it changes owner, flag, operator or name. Additionally, all fishing vessels longer than 15 metres are required to emit, every few minutes, signals containing (among other information) a timestamp, and their longitude and latitude.

Finding what vessel is conducting Illegal fishing activity is very hard to identify in this case, it is very hard to keep track of every vessel in the sea/ocean and finding whether they are conducting illegal fishing or not.

Over 68% of earth is ocean water and it is a lot of area to cover. Since there is a smaller number of eyes present in the ocean as compared to a busy road it is very rare to detect the illegal fishing activity physically, and to patrol these grounds it takes a lot of efforts, which is also a slower process. Hence, we need a way to know beforehand that where is the Illegal fishing activity going to happen, and that can be done by making predictions on the basis of Data Analysis, but since there is a large amount of scattered and unclassified data, it is hard to make those prediction.

3.4 DEPICTION OF PROBLEM STATEMENT

3.4.1 BLOCK DIAGRAM



3.5 OBJECTIVE

The main objective of this project is to predict the illegal fishing and to avoid them from happening. It works by considering the various attributes of the fishing vessels and their working hours and analysing then use the behavioural pattern to compare it with other fishing vessel activities and then classify them as illegal or non-illegal fishing. It also marks the country whose fishing vessel is classified as illegal fishing and then rank the countries respectively.

It also has the capability to use different type of algorithms to get the best possible outcome And then mark the countries in which there is a possibility of occurrence of illegal fishing

The main Objectives of the project are:

- Get the dataset
- Analyse the dataset for patterns
- Mark what is classified as illegal fishing and what is not.
- Test and run the model
- Blacklist the countries who are considered as IUU fishing promoters.

3.6 CONCLUSION

In this chapter we learnt about the need of this project and the problem it solves by getting replaced by the traditional model, hence increasing the effectiveness of the system. We took a look on the major problem faced by the current using model and the inefficiency in stopping the IUU fishing, and we looked at some data showing the need of this project in order for fulfilling the other needs of the society.

CHAPTER 4: METHODOLOGY

4.1 INTRODUCTION

Methodology involves algorithm and flowchart of how the control of the program flows upon execution, how the system was designed using various components, level by level working state of the model, and features of the model developed by our team. This chapter also gives you information about the Algorithms that we used in our system. All the information regarding the execution of the Model.

4.2 PROPOSED WORK

4.2.1 ALGORITHM

4.2.1.1 LOGISTIC REGRESSION

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import balanced_accuracy_score
train_X = train[num_col + bool_col]
train_y = train[target_col]
test_X = test[num_col + bool_col]
test_y = test[target_col]
train_y["illegal_fishing"]
type(train_y.values.ravel())
lr = LogisticRegression()
lr.fit(train_X, train_y.values.ravel())
predictions = lr.predict(test_X)
balanced_accuracy_score(test_y, predictions)
```

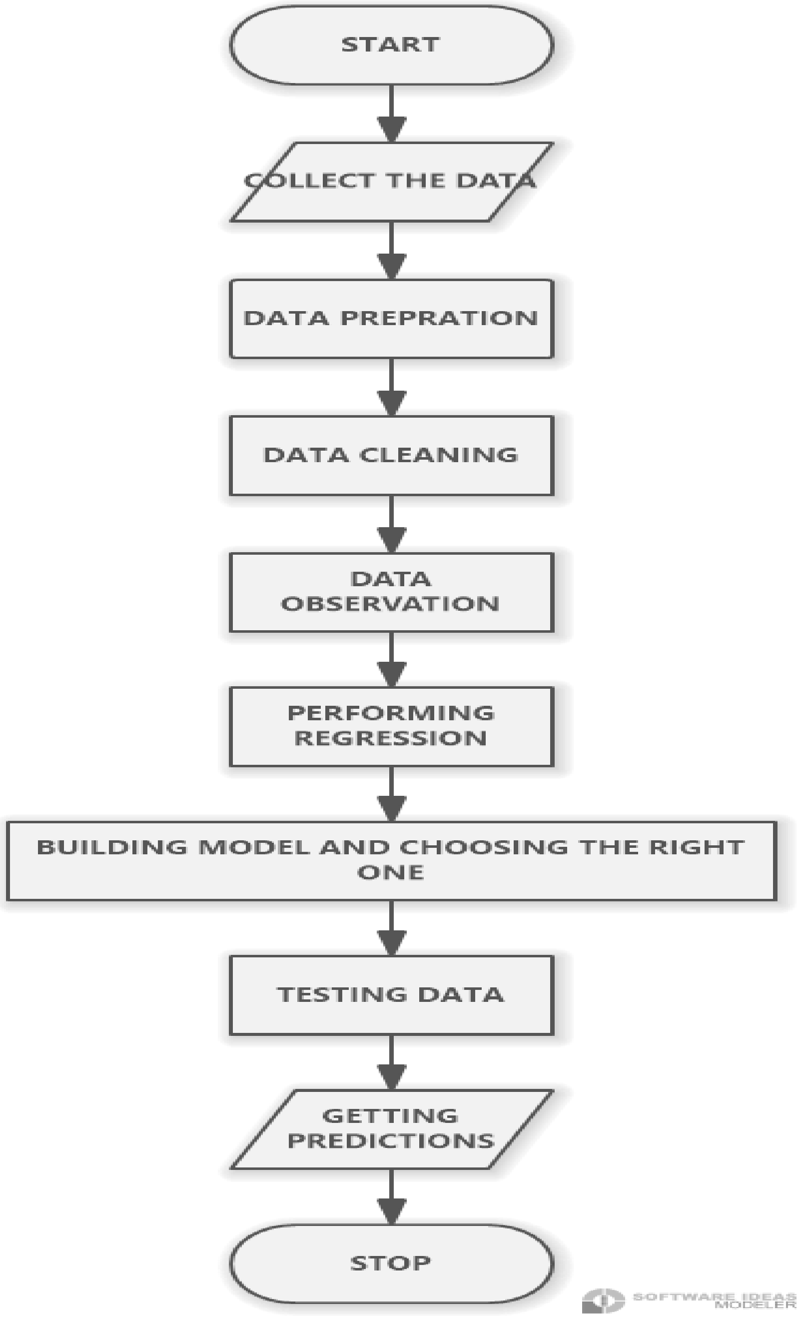
4.2.1.2 RANDOM FOREST

```
rf=RandomForestClassifier(random_state = 1,n_estimators=1000, n_jobs=-1)
rf.fit(train_X, train_y.values.ravel())
predictions = rf.predict(test_X)
balanced_accuracy_score(test_y, predictions)
predictions = clf.predict(test_X)
predictions
balanced_accuracy_score(test_y, predictions)
```

4.2.1.3 SQUID JIGGER ANALYSIS

```
squiddf=squiddf.loc[squiddf['gear_type_squid_jigger']==1]
squiddf=squiddf.drop(['gear_type_fixed_gear','gear_type_other_fishing','gear_t
ype_purse_seines','gear_type_trawlers'], axis=1)
train, test = train_test_split(squiddf, test_size=.25, random_state=42)
squiddf["illegal_fishing"].value_counts(normalize=True)
rf=RandomForestClassifier(random_state = 1,n_estimators=100, n_jobs=-1)
rf.fit(train_X, train_y.values.ravel())
predictions = rf.predict(test_X)
balanced_accuracy_score(test_y, predictions)
```

FLOW CHART:

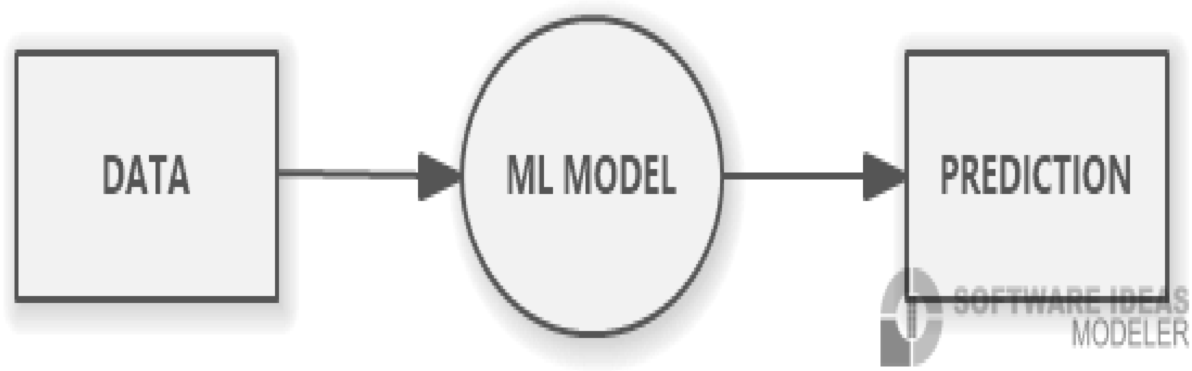


4.3 SYSTEM DESIGN

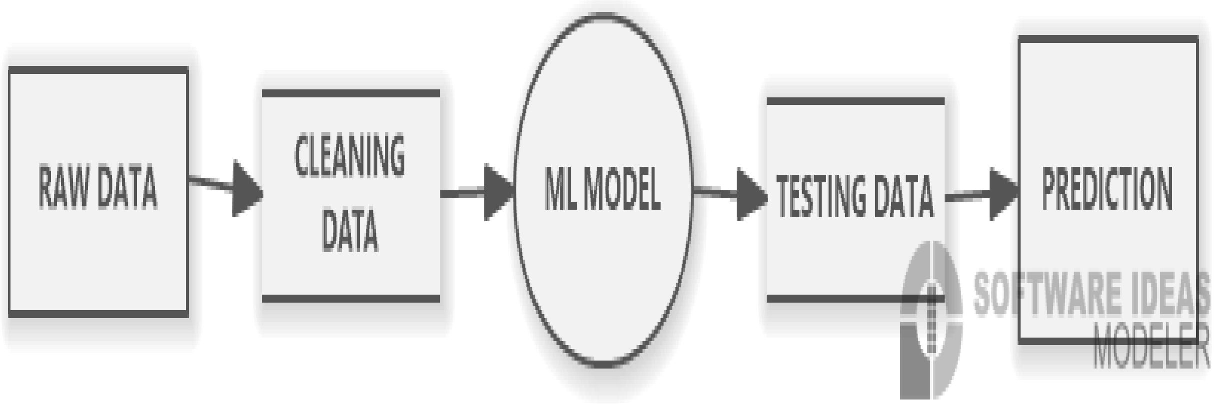
4.3.1 FUNCTIONAL SPECIFICATION OF SYSTEM

DATAFLOW DIAGRAM

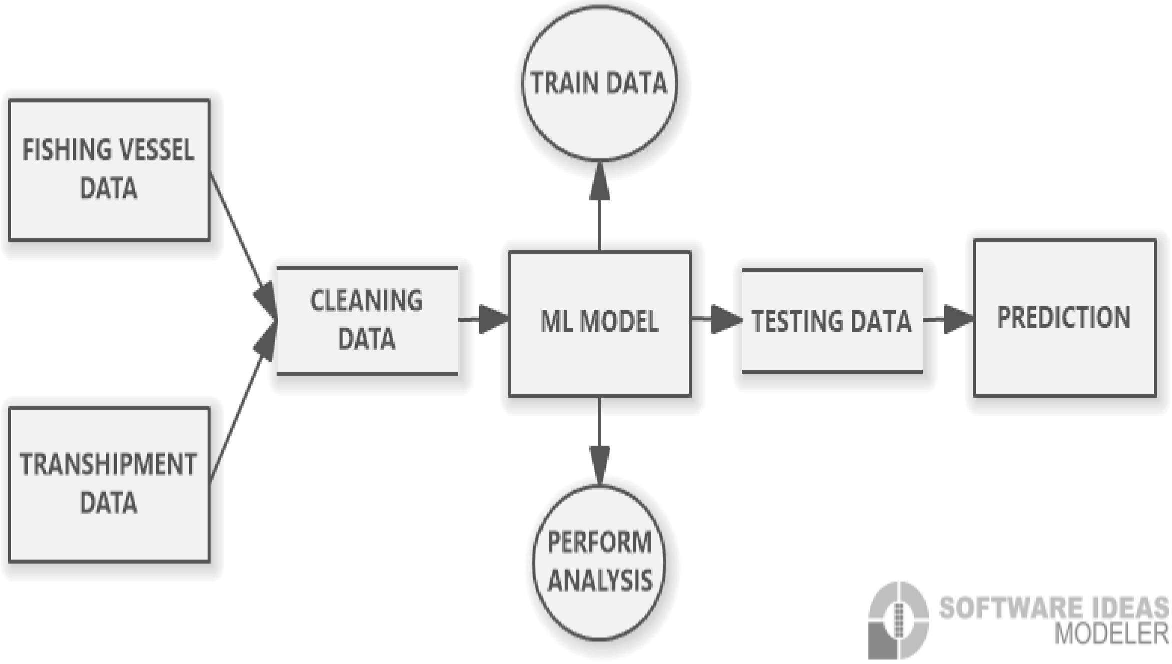
LEVEL 0:



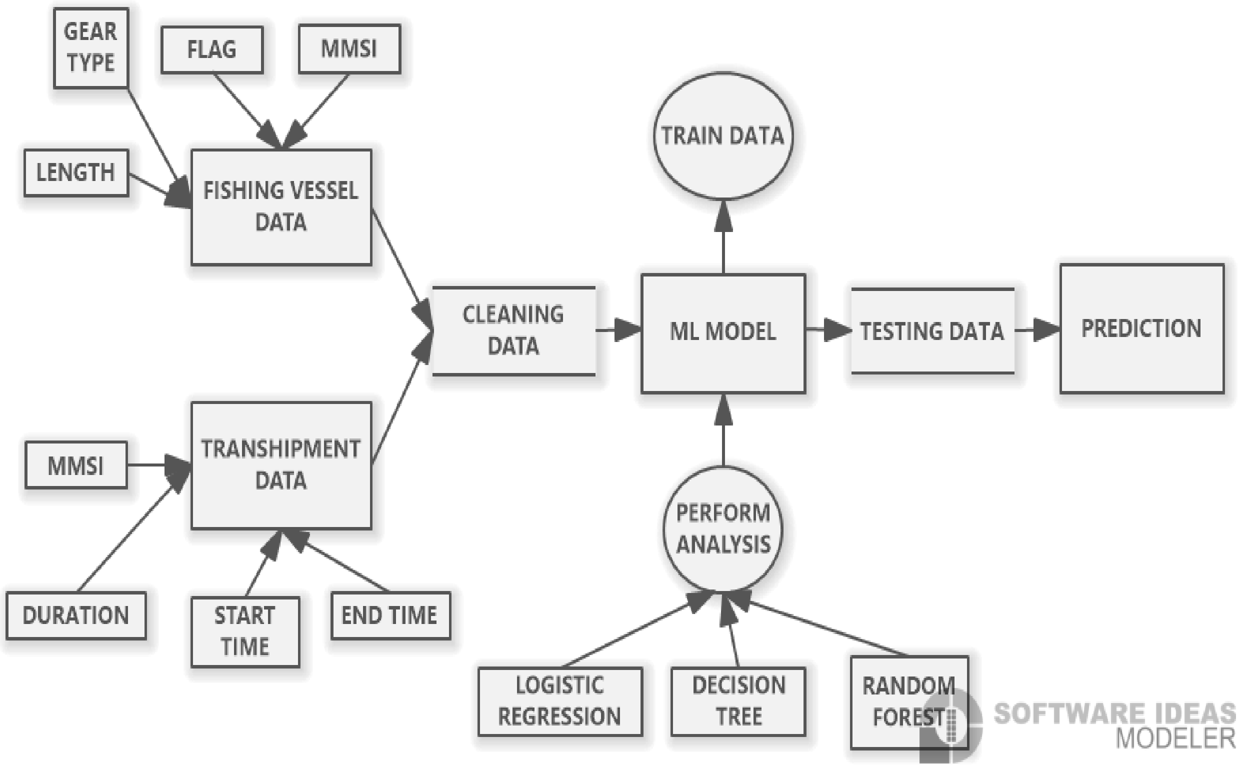
LEVEL 1:



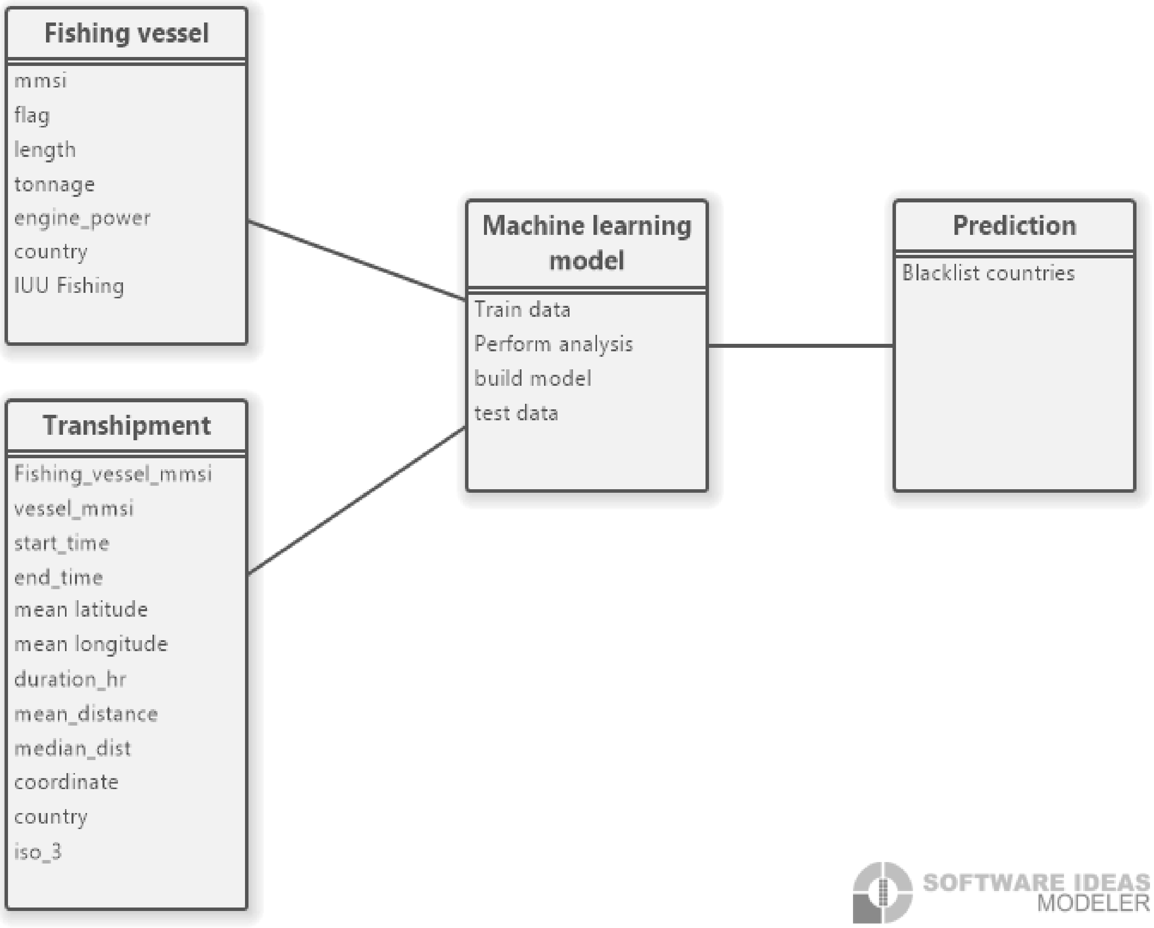
LEVEL 2:



LEVEL 3:



ENTITY RELATIONSHIP DIAGRAM



4.4 FEATURES

Get the data of fishing vessels and classify them as ‘very small’, ‘small’, ‘medium’, ‘large’, ‘very large’. And then based on this classification it performs the analysis by making a cut-off list for the vessel which are eligible for fishing operations, it also takes into the factor of ‘engine_power’ such that fishing vessels need a lot of engine power to carry the wight across the ports.

Fish vessels are also classified on the basis of tonnage i.e. the amount of weight it can carry or the payload it is authorised to carry. There are also classified as ‘very small’, ‘small’, ‘medium’, ‘large’, ‘very large’. Such that the vessels with medium and above are classifies as fishing vessels.

The duration hours fishing vessel spend in the ocean is also taken into consideration for the analysis of the illegal fishing concept, i.e., the amount of time spend in ocean by the vessel during fishing season and non-fishing seasons and based on those activities we are able to determine the overfishing rate of any vessel.

The model is also able to label the country flag i.e. it is able to tell which country vessels are performing the illegal fishing act and it is done through the help of the coordinates the ship were located at, so that we are able to know what countries are promoting illegal fishing or that who are not able to stop it.

4.5 MAJOR TOOLS USED IN MODEL

4.5.1 NUMPY:

NumPy forms the basis of powerful machine learning libraries like scikit-learn and SciPy. As machine learning grows, so does the list of libraries built on NumPy. TensorFlow’s deep learning capabilities have broad applications — among them speech and image recognition, text-based applications, time-series analysis, and video detection. PyTorch, another deep learning library, is popular among researchers in computer vision and natural language processing. MXNet is another AI package, providing blueprints and templates for deep learning.

Statistical techniques called ensemble methods such as binning, bagging, stacking, and boosting are among the ML algorithms implemented by tools such as XGBoost, LightGBM, and CatBoost — one of the fastest inference engines. Yellowbrick and Eli5 offer machine learning visualizations

4.5.2 SKLEARN

Scikit-learn is an open source machine learning library that supports supervised and unsupervised learning. It also provides various tools for model fitting, data preprocessing, model selection, model evaluation, and many other utilities

Scikit-learn provides dozens of built-in machine learning algorithms and models, called estimators. Each estimator can be fitted to some data using its fit method.

Here is a simple example where we fit a RandomForestClassifier to some very basic data:

```
from sklearn.ensemble import RandomForestClassifier
clf = RandomForestClassifier(random_state=0)
X = [[ 1,  2,  3],  # 2 samples, 3 features
     [11, 12, 13]]
 y = [0, 1]  # classes of each sample
clf.fit(X, y)
RandomForestClassifier(random_state=0)
```

The fit method generally accepts 2 inputs:

The samples matrix (or design matrix) X. The size of X is typically (n_samples, n_features), which means that samples are represented as rows and features are represented as columns

The target values y which are real numbers for regression tasks, or integers for classification (or any other discrete set of values). For t h ervised learning tasks, y does not need to be specified. Y is usually 1d array where the t h entry corresponds to the target of the t h sample (row) of X.

Both X and y are usually expected to be numpy arrays or equivalent array-like data types, though some estimators work with other formats such as sparse matrices.

Once the estimator is fitted, it can be used for predicting target values of new data

4.5.3 SEABORN

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures.

Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them

4.6 CONCLUSION

In this chapter, we looked at the major data flow of the circuit how to model is receiving its data and using that data to get the desired results, it is shown using flowcharts and DFD's, we also looked at some major tools used in this project in order for its proper functioning, such as, numpy, sklearn, seaborn. We also looked at the ER Diagram of the model which shows the working of project.

CHAPTER 5: IMPLEMENTATION

5.1 INTRODUCTION

This chapter deals with the discussion of how the model was designed and implemented. It involves the information related to working of the models and the result that are expected as outcome of this project. The pseudocode that is used for programming our processor used in our project is also included in this section.

5.2 DATA SET

Data set for the prediction of illegal fishing across all over the word has been taken from the research and development organization of Global Fish Watch Technology ,where the fishing is illegal or not is depending on different parameters such as :-

- Fishing vessels
- Transshipment
- Country and IUU rating

5.2.1 FISHING VESSEL

Data for the prediction is of around 6 years from 2012 to 2016 and the IUU rating is depend on different characterstics of Fishing vessels such as :

- MMSI
- FLAG
- GEARTYE
- LENGTH
- TONNAGE

- ENGINE POWER

	mmsi	flag	geartype	length	tonnage	engine_power	active_2012	active_2013	active_2014	active_2015	active_2016
0	603100157	AGO	trawlers	32.808468	299.003814	733.826977	False	False	False	True	True
1	603100137	AGO	trawlers	34.568782	395.683171	864.960188	False	False	False	True	True
2	603100161	AGO	trawlers	28.822140	263.849149	651.809642	False	False	False	True	True
3	603100174	AGO	trawlers	30.721429	299.700916	703.796086	False	False	False	True	True
4	603100164	AGO	trawlers	37.479248	405.967747	850.976640	False	False	False	True	True

```
RangeIndex(start=0, stop=73009, step=1)
```

5.2.2 TRANSSHIPMENT:

shipment_vessel_mmsi	start_time	end_time	mean_latitude	mean_longitude	duration_hr	median_distance_km	median_speed_knots	coordinates
354240000	2016-11-18T14:30:00Z	2016-11-19T01:50:00Z	-17.039085	-79.063725	11.333333	0.038188	0.585402	(-17.0390854214, -79.0637246039)
354240000	2016-12-11T14:50:00Z	2016-12-11T19:50:00Z	-20.269608	-79.244953	5.000000	0.020033	0.575663	(-20.26960835264285, -79.24495254404286)
354240000	2017-06-13T12:50:00Z	2017-06-15T01:20:00Z	-62.640767	-60.690240	36.500000	0.054992	0.019775	(-62.640767187979215, -60.69024019292565)
354240000	2016-11-15T11:30:00Z	2016-11-16T04:00:00Z	-17.046586	-79.061923	16.500000	0.036427	1.023917	(-17.04658591501366, -79.0619234553677)
354240000	2017-05-19T00:40:00Z	2017-05-19T20:50:00Z	-46.627878	-60.554922	20.166667	0.034053	0.544031	(-46.62787804705882, -60.55492183696471)

5.2.3 ILLEAGAL FISHING RATING:

	A	B	C
1	Country	IUU Fishing	
2	China	3.93	
3	Taiwan, Province of China	3.34	
4	Cambodia	3.23	
5	Russian Federation	3.16	
6	Vietnam	3.16	
7	Sierra Leone	3.01	
8	Yemen	2.96	
9	Sudan	2.77	
10	Liberia	2.76	
11	Somalia	2.75	
12	Libya	2.73	
13	Myanmar	2.73	
14	Mexico	2.71	
15	Philippines	2.71	
16	Indonesia	2.7	
17	Cameroon	2.69	
18	India	2.68	
19	Tanzania, United Republic o	2.65	
20	Japan	2.63	
21	Comoros	2.61	
22	Syria	2.61	
23	Timor-Leste	2.61	
24	Guinea	2.6	
25	Egypt	2.58	
26	Korea, Democratic People's	2.58	
27	Jamaica	2.57	
28	Panama	2.56	
29	Spain	2.56	

Tabelle1

Ready Accessibility: Good to go

5.2.4 COUNTRY _ISO:

	A	B	C	D
1	**Code**	**Name**		
2	ABW	Aruba		
3	AFG	Afghanistan		
4	AGO	Angola		
5	AIA	Anguilla		
6	ALA	Åland Islands		
7	ALB	Albania		
8	AND	Andorra		
9	ARE	United Arab Emirates		
10	ARG	Argentina		
11	ARM	Armenia		
12	ASM	American Samoa		
13	ATA	Antarctica		
14	ATF	French Southern Territories		
15	ATG	Antigua and Barbuda		
16	AUS	Australia		
17	AUT	Austria		
18	AZE	Azerbaijan		
19	BDI	Burundi		
20	BEL	Belgium		
21	BEN	Benin		
22	BES	Bonaire, Sint Eustatius and Saba		
23	BFA	Burkina Faso		
24	BGD	Bangladesh		
25	BGR	Bulgaria		
26	BHR	Bahrain		
27	BHS	Bahamas		
28	BIH	Bosnia and Herzegovina		
29	BLM	Saint Barthélemy		

Country_ISO

5.3 CLEANING OF DATA

Cleaning of data is done by removing the unnecessary data and keeping only the attributes which are needed for the prediction purpose. Hence we take the database and eliminate/drop the columns which are irrelevant or repeating. Resultant data we got after cleaning is as follows.

5.3.1 FISHING VESSELS:

	mmsi	flag	gear_type	length	tonnage	engine_power	country	IUU Fishing
0	603100157	AGO	trawlers	32.808468	299.003814	733.826977	Angola	2.37
1	603100137	AGO	trawlers	34.568782	395.683171	864.960188	Angola	2.37
2	603100161	AGO	trawlers	28.822140	263.849149	651.809642	Angola	2.37
3	603100174	AGO	trawlers	30.721429	299.700916	703.796086	Angola	2.37
4	603100164	AGO	trawlers	37.479248	405.967747	850.976640	Angola	2.37
5	603100139	AGO	trawlers	27.433138	277.298135	748.685601	Angola	2.37
6	603100175	AGO	trawlers	32.359115	442.479813	888.783860	Angola	2.37
7	603100159	AGO	trawlers	37.760114	443.995706	886.866416	Angola	2.37
8	603100160	AGO	trawlers	32.420990	395.693995	812.267999	Angola	2.37
9	603703900	AGO	purse_seines	36.787125	292.406624	1182.221917	Angola	2.37

5.3.2 TRANSSHIPMENT:

_vessel_mmsi	transshipment_vessel_mmsi	start_time	end_time	mean_latitude	mean_longitude	duration_hr	median_distance_km	median_speed_knots
416565000	354240000	2016-11-18T14:30:00Z	2016-11-19T01:50:00Z	-17.039085	-79.063725	11.333333	0.038188	0.585402
412679190	354240000	2016-12-11T14:50:00Z	2016-12-11T19:50:00Z	-20.269608	-79.244953	5.000000	0.020033	0.575663
440863000	354240000	2017-06-13T12:50:00Z	2017-06-15T01:20:00Z	-62.640767	-60.690240	36.500000	0.054992	0.019775
416563000	354240000	2016-11-15T11:30:00Z	2016-11-16T04:00:00Z	-17.046586	-79.061923	16.500000	0.036427	1.023917
441309000	354240000	2017-05-19T00:40:00Z	2017-05-19T20:50:00Z	-46.627878	-60.554922	20.166667	0.034053	0.544031

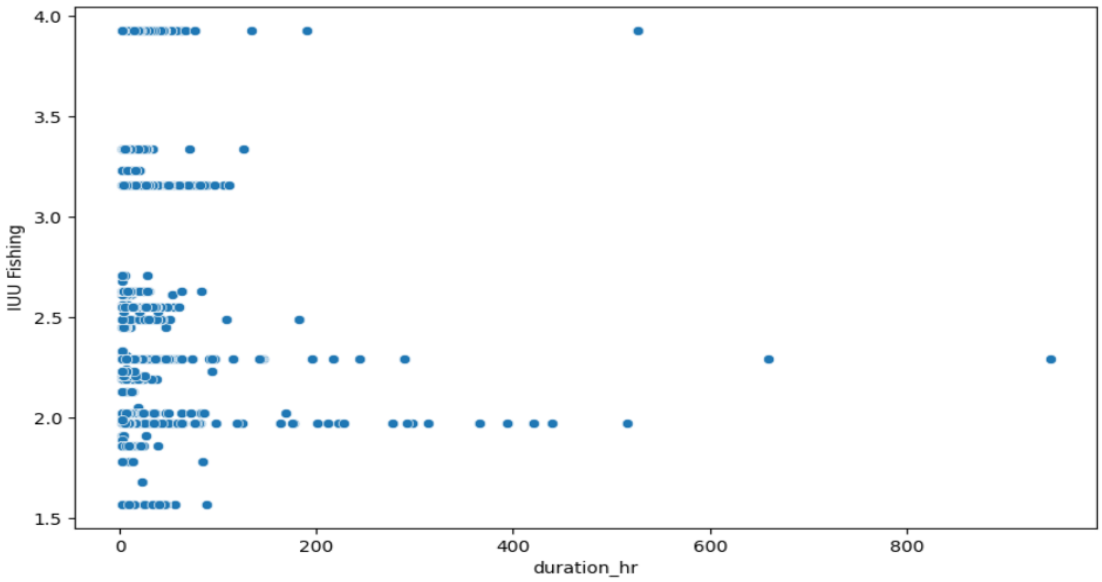
```
RangeIndex(start=0, stop=11681, step=1)
```

5.4 DATA PREPARATION:

In this stage we get the data ready for training purposes and observe certain patterns by plotting and observing the charts hoping to get a relation and pattern amongst the data attributes. We use several graphs like, scatterplot and bargraphs.

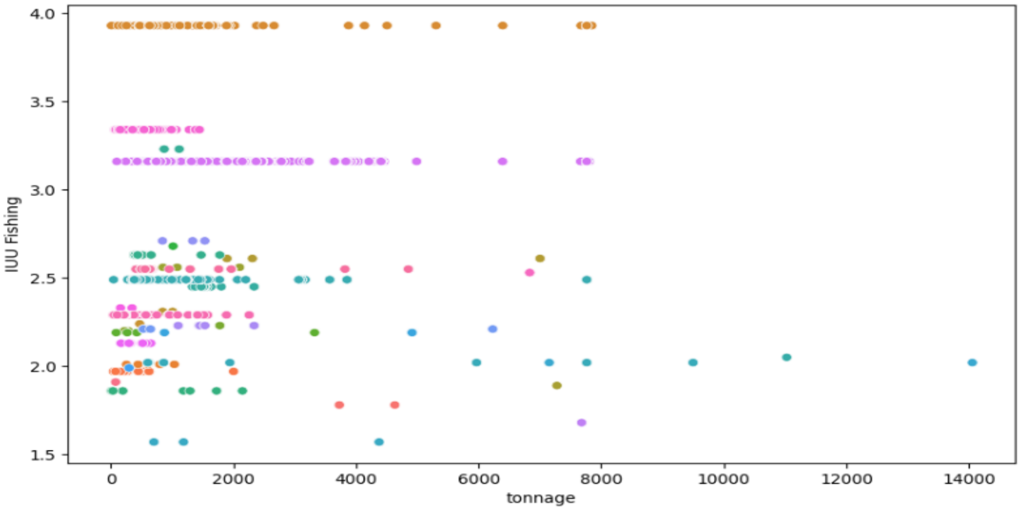
5.4.1 SCATTERPLOT

IUU FISHING VS DURATION_HR

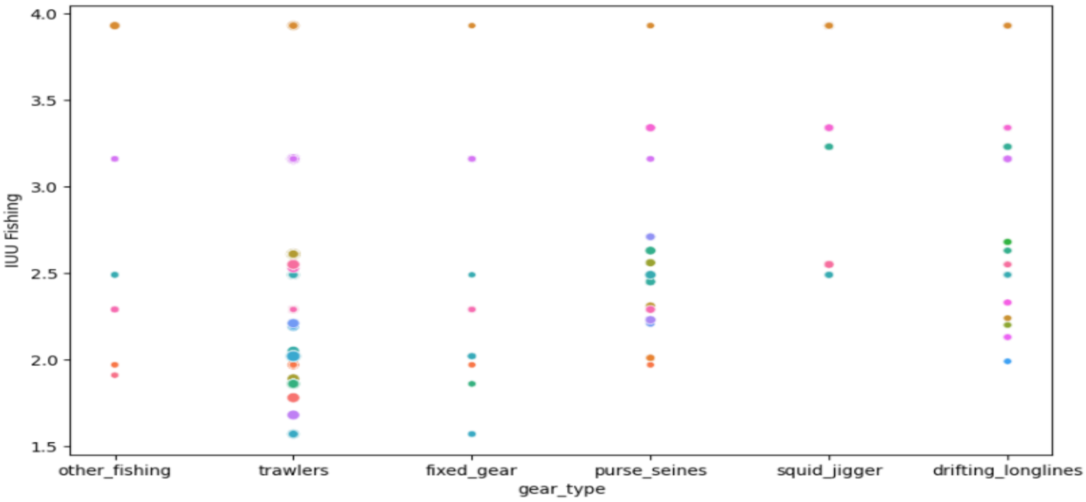


IUU FISHING VS LENGTH

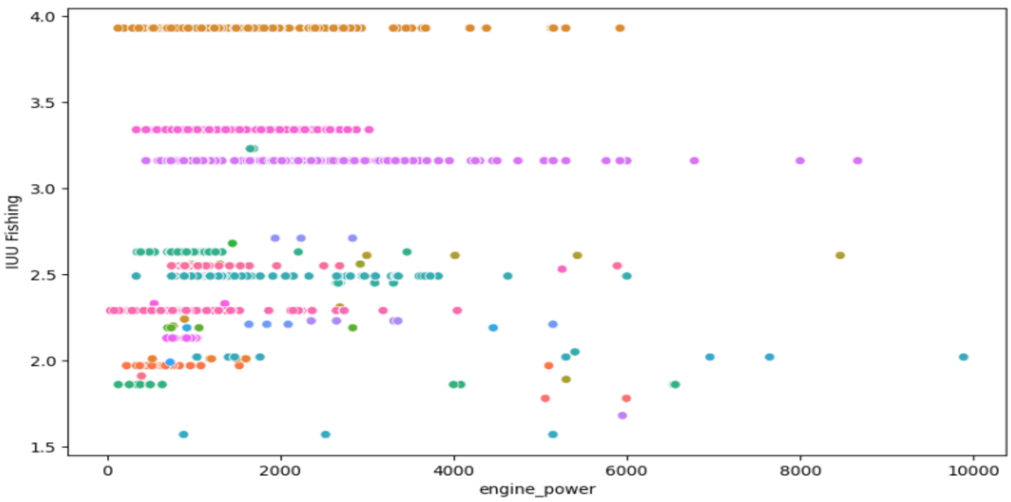
IUU FISHING VS TONNAGE



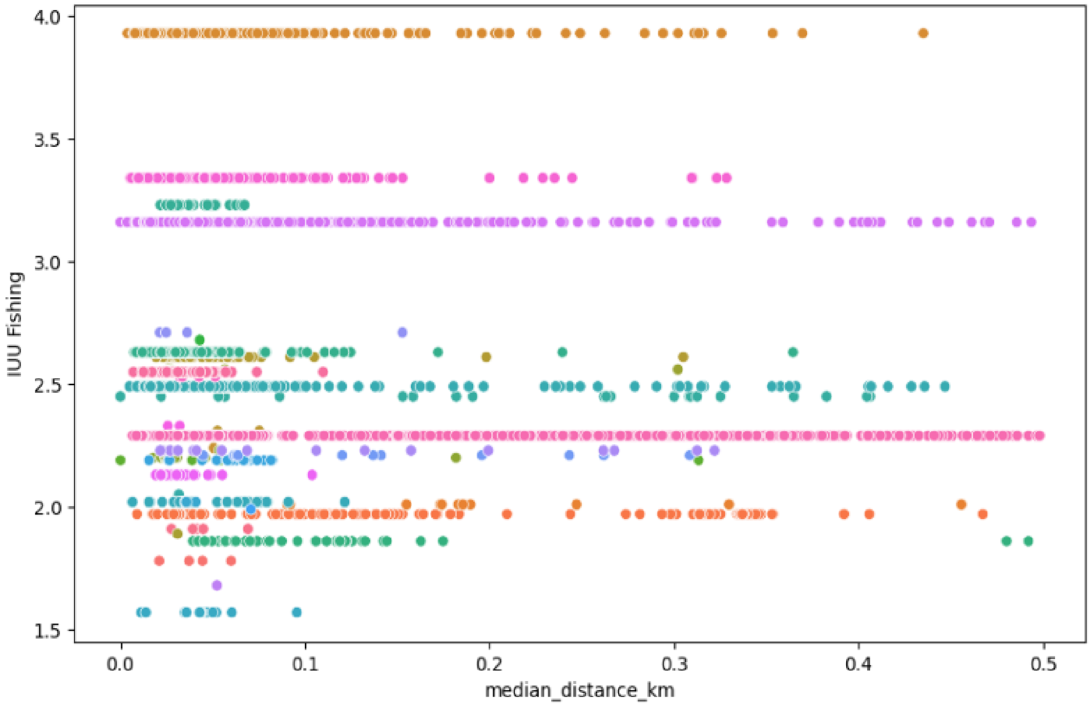
IUU FISHING VS GEAR_TYPE



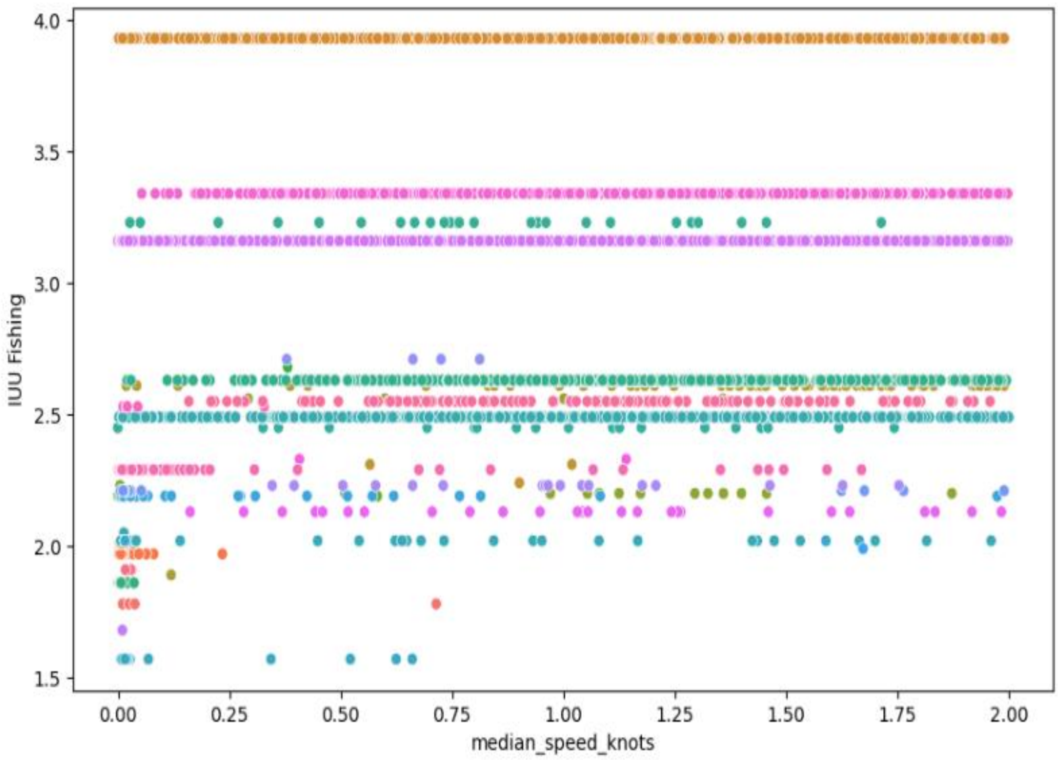
IUU FISHING VS ENGINE_POWER



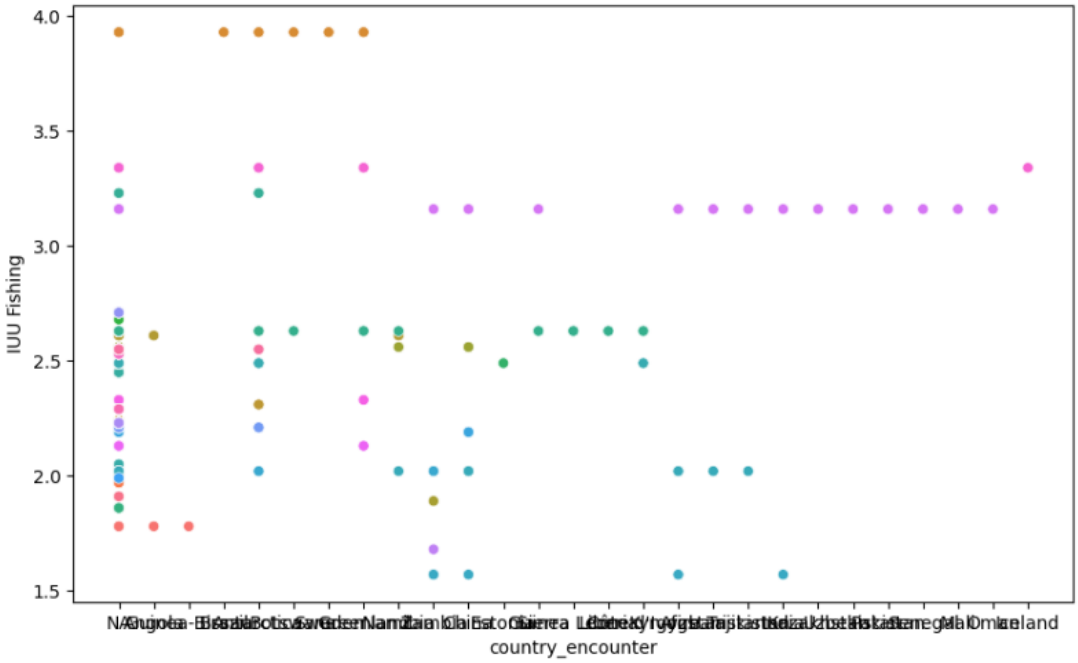
IUU FISHING VS MEDIAN DISTANCE



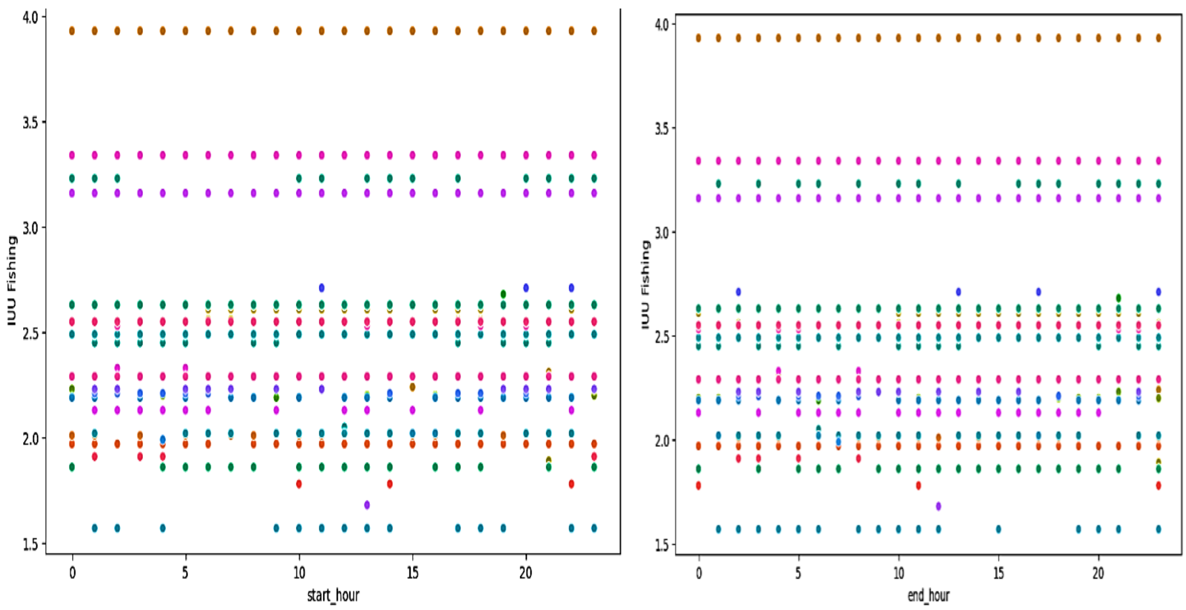
IUU FISHING VS MEDIAN SPEED



IUU FISHING VS COUNTRY ENCOUNTER



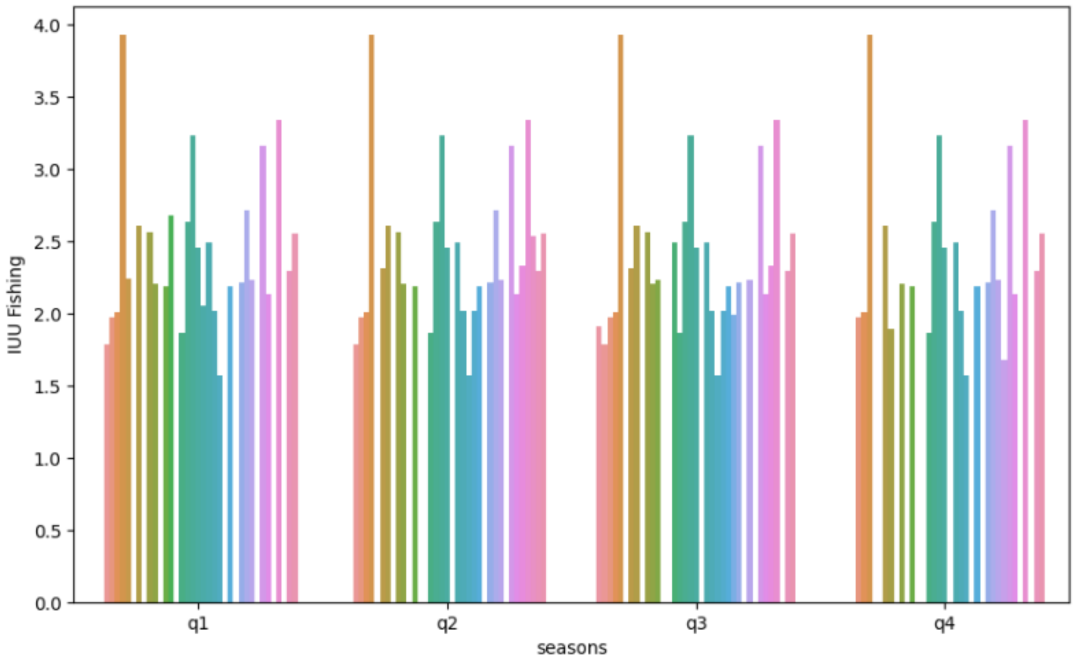
IUU FISHING VS START AND END HOUR



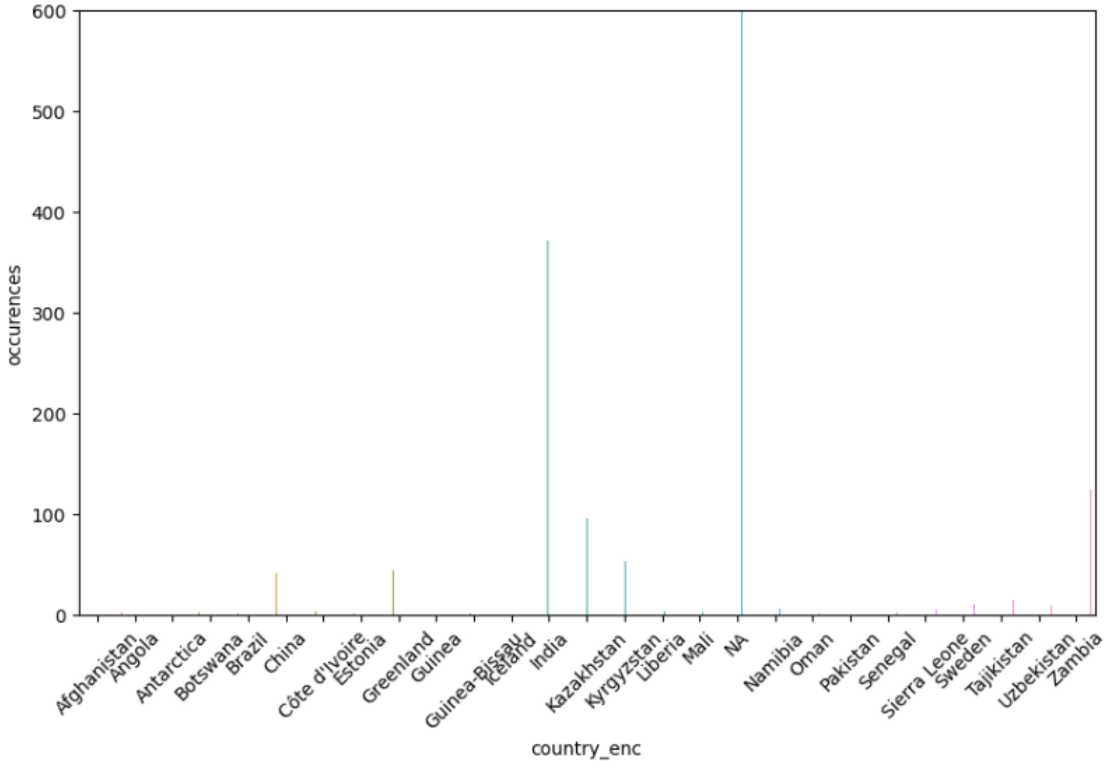
5.4.2 BARPLOT

Bar plots are done to observe the trends in the data.

IUU FISHING VS SEASONS



COUNTRY_ENC VS OCCURENCES



After observing and analyzing the patterns we worked on data and got the following results.

BEFORE:

	mmsi	flag	gear_type	length	tonnage	engine_power	country	IUU Fishing	fishing_vessel_mmsi	transshipment_vessel_mmsi	...	end_time	mean_l
0	603100157	AGO	trawlers	32.808468	299.003814	733.826977	Angola	2.37	NaN	NaN	...	NaN	
1	603100137	AGO	trawlers	34.568782	395.683171	864.960188	Angola	2.37	NaN	NaN	...	NaN	
2	603100161	AGO	trawlers	28.822140	263.849149	651.809642	Angola	2.37	NaN	NaN	...	NaN	
3	603100174	AGO	trawlers	30.721429	299.700916	703.796086	Angola	2.37	NaN	NaN	...	NaN	
4	603100164	AGO	trawlers	37.479248	405.967747	850.976640	Angola	2.37	NaN	NaN	...	NaN	

5 rows × 21 columns

AFTER:

	mmsi	gear_type	length	tonnage	engine_power	country	IUU Fishing	illegal_fishing
0	603100157	trawlers	32.808468	299.003814	733.826977	Angola	2.37	0
1	603100137	trawlers	34.568782	395.683171	864.960188	Angola	2.37	0
2	603100161	trawlers	28.822140	263.849149	651.809642	Angola	2.37	0
3	603100174	trawlers	30.721429	299.700916	703.796086	Angola	2.37	0
4	603100164	trawlers	37.479248	405.967747	850.976640	Angola	2.37	0

5.5 DATA ANALYSIS

5.5.1 OLS REGRESSION

OLS Regression Results

Dep. Variable:	IUU Fishing	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	1.290e+07
Date:	Sun, 13 Nov 2022	Prob (F-statistic):	0.00
Time:	12:35:52	Log-Likelihood:	48108.
No. Observations:	10530	AIC:	-9.614e+04
Df Residuals:	10493	BIC:	-9.587e+04
Df Model:	36		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.0240	5.91e+08	5.11e-09	1.000	-1.16e+09	1.16e+09
Australia	-1.1158	5.91e+08	-1.89e-09	1.000	-1.16e+09	1.16e+09
Belize	-1.2463	5.91e+08	-2.11e-09	1.000	-1.16e+09	1.16e+09
Cambodia	0.2053	5.91e+08	3.47e-10	1.000	-1.16e+09	1.16e+09
Canada	-1.0559	5.91e+08	-1.79e-09	1.000	-1.16e+09	1.16e+09
Chile	-1.0163	5.91e+08	-1.72e-09	1.000	-1.16e+09	1.16e+09
China	0.9015	5.91e+08	1.52e-09	1.000	-1.16e+09	1.16e+09
Colombia	-0.7157	5.91e+08	-1.21e-09	1.000	-1.16e+09	1.16e+09
Comoros	-0.4141	5.91e+08	-7e-10	1.000	-1.16e+09	1.16e+09
Côte d'Ivoire	-0.7842	5.91e+08	-1.33e-09	1.000	-1.16e+09	1.16e+09
Fiji	-0.8250	5.91e+08	-1.4e-09	1.000	-1.16e+09	1.16e+09
Germany	-1.1342	5.91e+08	-1.92e-09	1.000	-1.16e+09	1.16e+09
Iceland	-1.1660	5.91e+08	-1.97e-09	1.000	-1.16e+09	1.16e+09
India	-0.3446	5.91e+08	-5.83e-10	1.000	-1.16e+09	1.16e+09
Iran, Islamic Republic of	-0.5350	5.91e+08	-9.05e-10	1.000	-1.16e+09	1.16e+09
Japan	-0.3949	5.91e+08	-6.68e-10	1.000	-1.16e+09	1.16e+09
Kiribati	-0.5757	5.91e+08	-9.74e-10	1.000	-1.16e+09	1.16e+09
Korea, Republic of	-0.5354	5.91e+08	-9.05e-10	1.000	-1.16e+09	1.16e+09
Latvia	-1.4548	5.91e+08	-2.46e-09	1.000	-1.16e+09	1.16e+09
Lithuania	-1.0044	5.91e+08	-1.7e-09	1.000	-1.16e+09	1.16e+09
Micronesia, Federated States of	-0.7944	5.91e+08	-1.34e-09	1.000	-1.16e+09	1.16e+09
Netherlands	-1.0043	5.91e+08	-1.7e-09	1.000	-1.16e+09	1.16e+09
Norway	-0.8343	5.91e+08	-1.41e-09	1.000	-1.16e+09	1.16e+09
Oman	-1.0353	5.91e+08	-1.75e-09	1.000	-1.16e+09	1.16e+09
Papua New Guinea	-0.7956	5.91e+08	-1.35e-09	1.000	-1.16e+09	1.16e+09
Peru	-0.8156	5.91e+08	-1.38e-09	1.000	-1.16e+09	1.16e+09
Philippines	-0.3160	5.91e+08	-5.34e-10	1.000	-1.16e+09	1.16e+09
Poland	-1.3461	5.91e+08	-2.28e-09	1.000	-1.16e+09	1.16e+09
Russian Federation	0.1340	5.91e+08	2.27e-10	1.000	-1.16e+09	1.16e+09
Saint Kitts and Nevis	-0.9741	5.91e+08	-1.65e-09	1.000	-1.16e+09	1.16e+09
Seychelles	-0.8943	5.91e+08	-1.51e-09	1.000	-1.16e+09	1.16e+09
Spain	-0.4659	5.91e+08	-7.88e-10	1.000	-1.16e+09	1.16e+09
Taiwan, Province of China	0.3141	5.91e+08	5.31e-10	1.000	-1.16e+09	1.16e+09
Thailand	-0.6956	5.91e+08	-1.18e-09	1.000	-1.16e+09	1.16e+09
Ukraine	-0.4942	5.91e+08	-8.36e-10	1.000	-1.16e+09	1.16e+09
United Kingdom	-0.8354	5.91e+08	-1.41e-09	1.000	-1.16e+09	1.16e+09
United States	-0.7315	5.91e+08	-1.24e-09	1.000	-1.16e+09	1.16e+09
Vanuatu	-0.4715	5.91e+08	-7.98e-10	1.000	-1.16e+09	1.16e+09

Omnibus:	1286.947	Durbin-Watson:	0.001
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1918.200
Skew:	-0.902	Prob(JB):	0.00
Kurtosis:	4.058	Cond. No.	1.69e+14

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 4.76e-25. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.

5.5.2 LINEAR MODEL (IUU AND COUNTRY)

	const	gear_type	length	tonnage	engine_power	IUU Fishing	duration_hr	median_distance_km	median_speed_knots	country_encounter	...	Russian Federation
0	1.0	other_fishing	23.009784	78.895456	394.867663	1.91	2.166667	0.069157	0.031628	NA	...	0
1	1.0	other_fishing	23.009784	78.895456	394.867663	1.91	2.000000	0.027717	0.024307	NA	...	0
2	1.0	other_fishing	23.009784	78.895456	394.867663	1.91	2.000000	0.041409	0.018666	NA	...	0
3	1.0	other_fishing	23.009784	78.895456	394.867663	1.91	3.833333	0.044934	0.029034	NA	...	0
4	1.0	other_fishing	23.009784	78.895456	394.867663	1.91	26.333333	0.039192	0.016855	NA	...	0

5 rows × 52 columns

```
lindf_2.corr()['IUU Fishing']
```

```
const                              NaN
length                        0.251650
tonnage                       0.141647
engine_power                  0.151534
IUU Fishing                   1.000000
duration_hr                  -0.082865
median_distance_km           -0.312466
median_speed_knots            0.243235
start_month                   0.010185
start_hour                    0.045726
end_month                     0.008597
end_hour                      0.014650
Australia                    -0.048152
Belize                       -0.047866
Cambodia                      0.014199
Canada                       -0.278689
Chile                        -0.065316
China                         0.713218
Colombia                     -0.020007
Comoros                      -0.069074
Côte d'Ivoire                -0.015438
Fiji                         -0.060565
Germany                      -0.021899
Iceland                      -0.136811
India                        -0.007316
Iran, Islamic Republic of    -0.010824
Japan                        -0.205603
Kiribati                     -0.053034
Korea, Republic of           -0.319102
Latvia                       -0.124466
Lithuania                    -0.115547
Micronesia, Federated States of  -0.015623
```

```
Netherlands                  -0.033777
Norway                       -0.083526
Oman                         -0.020053
Papua New Guinea             -0.064464
Peru                         -0.053065
Philippines                  -0.013527
Poland                       -0.025776
Russian Federation            0.143936
Saint Kitts and Nevis        -0.018946
Seychelles                   -0.090884
Spain                        -0.019065
Taiwan, Province of China     0.157269
Thailand                     -0.019485
Ukraine                      -0.022555
United Kingdom               -0.032727
United States                -0.495432
Vanuatu                      -0.112759
Name: IUU Fishing, dtype: float64
```

5.2.3 LOGISTIC REGRESSION

```
lr = LogisticRegression()
lr.fit(train_X, train_y.values.ravel())
```

```
▾ LogisticRegression
LogisticRegression()
```

```
predictions = lr.predict(test_X)
```

```
balanced_accuracy_score(test_y, predictions)
```

```
0.793511123487442
```

```
from sklearn.metrics import classification_report
print(classification_report(test_y, predictions))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99     16126
           1       0.68      0.59      0.63       424

    accuracy                           0.98     16550
   macro avg       0.84      0.79      0.81     16550
weighted avg       0.98      0.98      0.98     16550
```

5.2.4 RANDOM FOREST

```
rf=RandomForestClassifier(random_state = 1,n_estimators=1000, n_jobs=-1)
rf.fit(train_X, train_y.values.ravel())
```

```
▾ RandomForestClassifier
RandomForestClassifier(n_estimators=1000, n_jobs=-1, random_state=1)
```

```
predictions = rf.predict(test_X)
balanced_accuracy_score(test_y, predictions)
```

```
0.8177481753362085
```

```
predictions
```

```
array([0, 0, 0, ..., 0, 0, 0])
```

```
test_X
```

	IUU Fishing	gear_type_fixed_gear	gear_type_other_fishing	gear_type_purse_seines	gear_type_squid_jigger	gear_type_trawlers	Albania	Algeria	Angola
24680	3.93	0	0	0	0	1	0	0	0
70765	3.34	0	0	0	0	0	0	0	0
43735	3.93	0	1	0	0	0	0	0	0
7546	3.93	0	0	0	0	1	0	0	0
16903	3.93	0	0	0	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...
54722	2.50	0	0	0	0	1	0	0	0
36603	3.93	1	0	0	0	0	0	0	0
27229	3.93	0	0	0	0	1	0	0	0
2080	3.93	0	0	0	0	1	0	0	0
57404	2.49	0	0	0	0	1	0	0	0

16550 rows × 127 columns

```
print(classification_report(test_y, predictions))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99     16126
           1       0.67      0.64      0.66       424

    accuracy                           0.98     16550
   macro avg       0.83      0.82      0.82     16550
weighted avg       0.98      0.98      0.98     16550
```

5.2.5 GRADIENTBOOSTINGCLASSIFIER

```
clf = GradientBoostingClassifier(random_state=0)
clf.fit(train_X, train_y.values.ravel())
```

```
GradientBoostingClassifier
GradientBoostingClassifier(random_state=0)
```

```
predictions = clf.predict(test_X)
```

```
balanced_accuracy_score(test_y, predictions)
```

```
0.7684979021339031
```

```
print(classification_report(test_y, predictions))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99     16126
           1       0.72      0.54      0.62       424

    accuracy                           0.98     16550
   macro avg       0.86      0.77      0.81     16550
weighted avg       0.98      0.98      0.98     16550
```

5.2.6 CROSS VALIDATING

```
Model: Logistic Regression
Mean Balanced Accuracy Score: 0.6985079476374081


Model: Decision Tree
Mean Balanced Accuracy Score: 0.7920357875930553


Model: Random Forest
Mean Balanced Accuracy Score: 0.8042728312376897
```

Hence, we can see, Random Forest is the best suitable algorithm as compared to Logistic regression and Decision tree in this case.

5.6 MODEL BUILDING

MACHINE LEARNING PREPRATION

5.6.1 PREPARING A PIVOT TABLE

Creating a pivot table to identify/Predicting the illegal fishing activities.

id_encounter	9	10	11	12	13	14	15	16	17	18	...	10520	10521	10522	10523	10524	10525	10526	10527	10528	10529
country_id																					
2	1.97	1.97	1.97	1.97	1.97	1.97	1.97	1.97	1.97	1.97	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
36	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2.55	2.55	2.55	2.55	2.55	2.55	2.55	2.55	2.55	2.55

20 rows × 10489 columns

After creating the pivot table now, we train the model to fill it with the appropriate data using certain algorithms and then test the same for the efficient results.

```
#import more libraries

import matplotlib.pyplot as plt
import seaborn as sns
from scipy.sparse import csr_matrix
from sklearn.neighbors import NearestNeighbors

%matplotlib inline
```

```
country_features = csr_matrix(features)
```

```
#train the model
```

```
model_knn = NearestNeighbors(metric="cosine",
                             algorithm="brute",
                             n_jobs=-1)
model_knn.fit(country_features)
```

```
NearestNeighbors
NearestNeighbors(algorithm='brute', metric='cosine', n_jobs=-1)
```

Table after applying algorithm:

id_encounter	9	10	11	12	13	14	15	16	17	18	...	10520	10521	10522	10523	10524	10525	10526	10527	10528	10529
country_id																					
2	1.97	1.97	1.97	1.97	1.97	1.97	1.97	1.97	1.97	1.97	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	2.55	2.55	2.55	2.55	2.55	2.55	2.55	2.55	2.55	2.55

20 rows × 10489 columns

Country ID setup

		IUU Fishing	id_encounter	nr_occurrences
country	country_id			
Australia	0	5	5	5
Belize	1	4	4	4
Cambodia	17	25	25	25
Canada	2	183	183	183
Chile	3	11	11	11
China	4	1715	1715	1715
Colombia	6	2	2	2
Comoros	7	64	64	64
Côte d'Ivoire	5	1	1	1
Fiji	10	14	14	14
Germany	8	1	1	1
Iceland	15	37	37	37
India	13	1	1	1
Iran, Islamic Republic of	14	1	1	1
Japan	16	588	588	588
Kiribati	18	21	21	21
Korea, Republic of	20	803	803	803
Latvia	22	20	20	20
Lithuania	21	35	35	35
Micronesia, Federated States of	11	1	1	1
Netherlands	23	3	3	3
Norway	24	26	26	26
Oman	25	1	1	1
Papua New Guinea	28	17	17	17
Peru	26	11	11	11
Philippines	27	4	4	4
Poland	29	1	1	1
Russian Federation	30	4760	4760	4760
Saint Kitts and Nevis	19	1	1	1
Seychelles	31	27	27	27
Spain	9	4	4	4
Taiwan, Province of China	33	950	950	950
Thailand	32	2	2	2
Ukraine	34	5	5	5
United Kingdom	12	4	4	4
United States	35	1049	1049	1049
Vanuatu	36	133	133	133

FINAL PREDICTION LIST

```python
#creating a data frame to take the right country_id and match with the index

newc_list=totaldf2.groupby('country', as_index=False).agg({'country_id':'mean'}).sort_values(by='country_id', ascending=True)
newc_list
```

	country	country_id
1	Canada	2.0
2	Chile	3.0
3	China	4.0
4	Comoros	7.0
5	Fiji	10.0
6	Iceland	15.0
7	Japan	16.0
0	Cambodia	17.0
8	Kiribati	18.0
9	Korea, Republic of	20.0
11	Lithuania	21.0
10	Latvia	22.0
12	Norway	24.0
14	Peru	26.0
13	Papua New Guinea	28.0
15	Russian Federation	30.0
16	Seychelles	31.0
17	Taiwan, Province of China	33.0
18	United States	35.0
19	Vanuatu	36.0

```python
# taking the indeces

distances, indeces = model_knn.kneighbors(features.loc[4].values.reshape(1, -1), n_neighbors=5)
```

```python
#confirming indeces

indeces
```

```
array([[ 2, 13, 15, 17, 14]], dtype=int64)
```

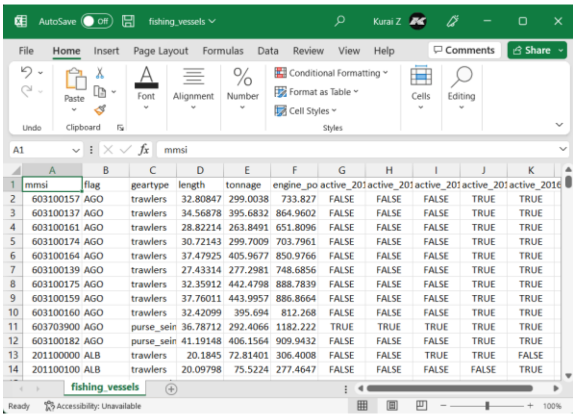
5.7 CONCLUSION

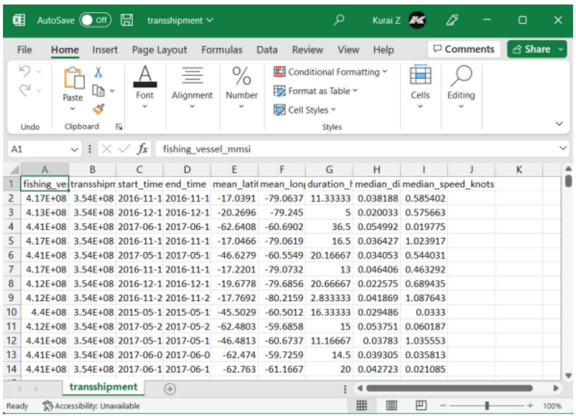
This chapter was all about the implementation of the circuit how the Model is made and how it gets the required predictions to operate properly, it includes the clear implementation of the model implemented, it also shows the proper application of the data gathered in order for working in the desired manner, we also showed the proper working of the model how the predictions are being made based on the algorithms.

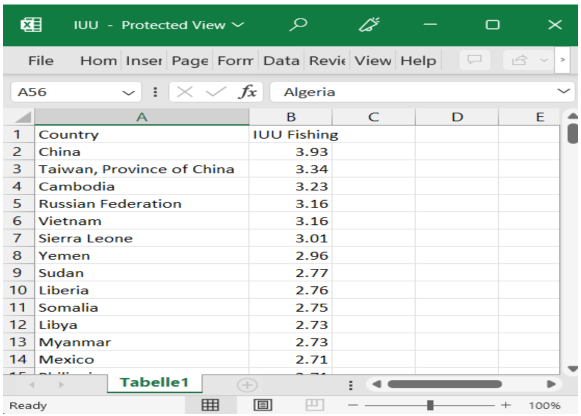
CHAPTER 6: RESULTS AND OBSERVATION

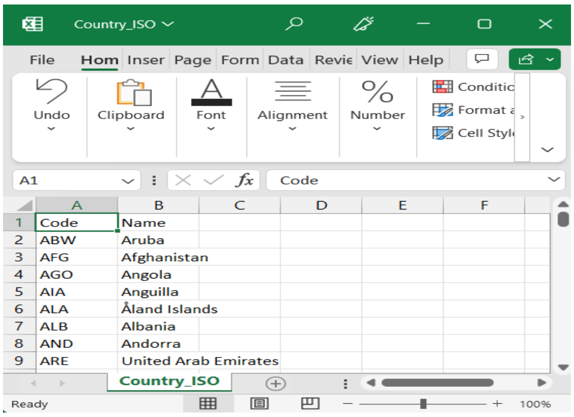
We have obtained unique results at the different stages of the model building.

[1] Our first step was to gather the right amount of data in this vast ocean of raw data available, in order to move further.

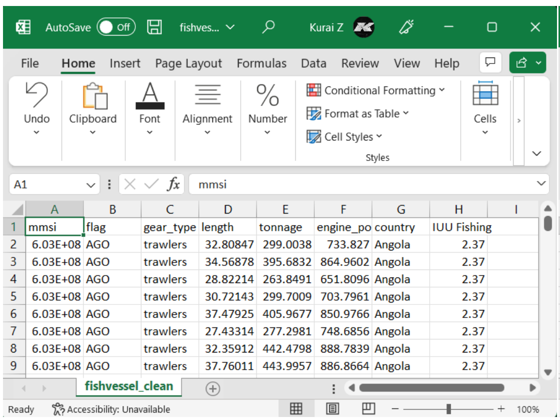


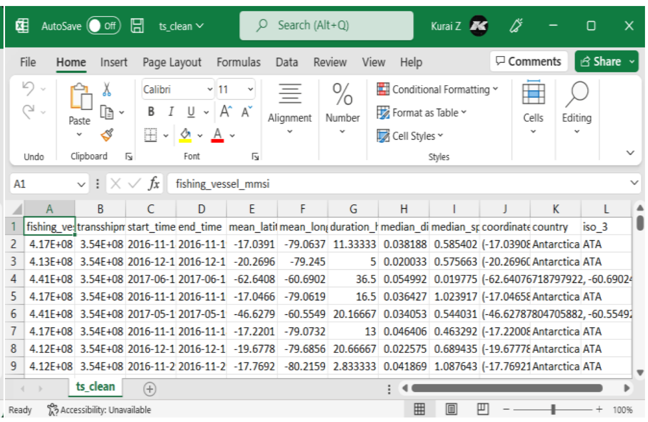




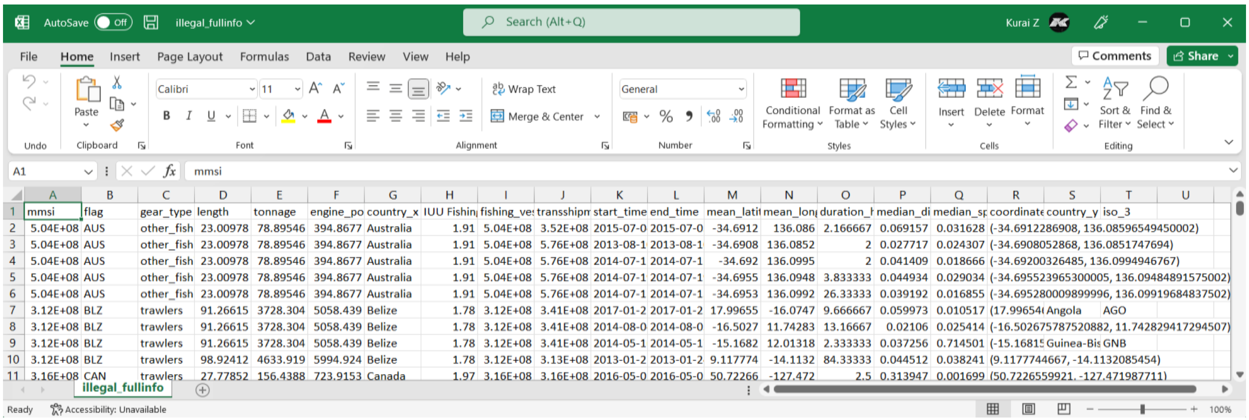


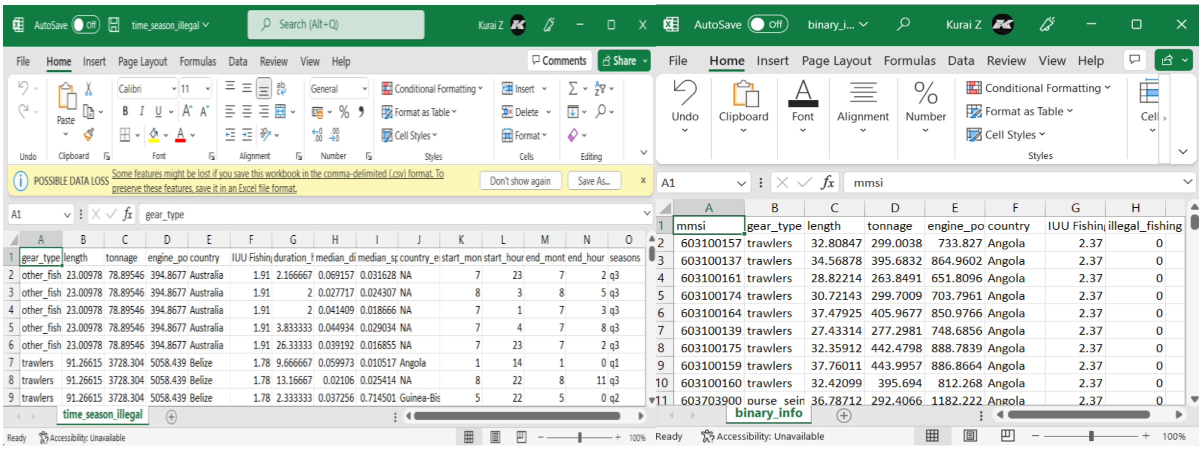
[2] Second step was to clean this data for the analysis purposes so that only those data remain who will help in the illegal fishing detection





[3] After cleaning of data next step was to perform the data analysis and group together the relevant data so that our prediction model can have much better accuracy.





[4] Based on the following data and algorithms our model predicted the following results and it also named some countries which shows the sign of increasing illegal fishing.

	country	country_id
1	Canada	2.0
2	Chile	3.0
3	China	4.0
4	Comoros	7.0
5	Fiji	10.0
6	Iceland	15.0
7	Japan	16.0
0	Cambodia	17.0
8	Kiribati	18.0
9	Korea, Republic of	20.0
11	Lithuania	21.0
10	Latvia	22.0
12	Norway	24.0
14	Peru	26.0
13	Papua New Guinea	28.0
15	Russian Federation	30.0
16	Seychelles	31.0
17	Taiwan, Province of China	33.0
18	United States	35.0
19	Vanuatu	36.0

	country	IUU Fishing	id_encounter	nr_occurrences	country_id
9	Canada	1.97	9	183	2
10	Canada	1.97	10	183	2
11	Canada	1.97	11	183	2
12	Canada	1.97	12	183	2
13	Canada	1.97	13	183	2

BLACKLSITED COUNTRIES (BASED ON CHINEASE ILLEGAL FISHING BEHAVIOUR)

```
BLACKLIST
Countries with similar behaviour to China


1:  Peru
2:  Russian Federation
3:  Taiwan, Province of China
4:  Papua New Guinea
```

This result was based on China’s illegal fishing behaviors and we used KNN- neighbor algo to identify similar activities, we choose China because our model already predicted that China was already having very large amount of IUU Fishing going on, hence we took it as reference. We can also do the same with other countries by taking them as reference but they should have a higher IUU fishing rate in order to get the right prediction.

CHAPTER 7: PROJECT RELEVANCE

7.1 INTRODUCTION

This chapter walks you through the facts that clearly state how our project which is "Illegal fishing prediction system" is going to contribute to the society in terms of reducing the illegal fishing activities and also reducing the manual efforts by making them predicting them beforehand. Going through this section will also inform you about the technical novelty and utility of the project, objective and relevance of the project and the expected outcome of the project.

7.2 CONTRIBUTION OF PROJECT TO THE SOCIETY:

IUU fishing is key to improving the sustainability of fisheries around the world, and to secure the health of our oceans now and in the future. Important secondary outcomes are diminished geopolitical tension and economic impacts on local economies. Current IUU monitoring relies on detecting spoofing in vessel location records, which is retrospective and confined to identifying anomalous spatial behaviour of vessels one at a time. Both are useful but limited in their ability to inform enforcement agencies like coast guards about the occurrence of IUU activity.

To overcome these challenges, we have developed an approach to IUU detection, and even prediction. This is a promising new tool for detecting, predicting and deterring IUU fishing. To find suspected illegal fishers, national authorities have long relied on conventional maritime patrols, which are costly, inefficient, often dangerous, and largely ineffective.

Many of the world's most important fisheries are experiencing increases in illegal fishing, undermining efforts to sustainably conserve and manage fish stocks. A major challenge to ending illegal, unreported, and unregulated (IUU) fishing is improving our ability to identify whether a vessel is fishing illegally and where illegal fishing is likely to occur in the ocean. However, monitoring the oceans is costly, time-consuming, and logistically challenging for maritime authorities to patrol. To address this problem, we use vessel tracking data and machine learning to predict illegal fishing, we focus on Chinese fishing vessels, which have consistently fished illegally in this region. We combine vessel location data with oceanographic seascapes -- classes of oceanic areas based on oceanographic variables -- as well as other remotely sensed oceanographic variables to train a series of machine learning models of varying levels of complexity. Hence it is efficiently able to predict the illegal fishing activity and we are able to stop those activities by the help of law enforcement. Resulting in saving the marine life from excess fishing and saving the bio-diversity of earth maintaining the proper balance.

7.3 OBJECTIVE AND RELEVANCE OF PROJECT:

The main objective of this project is to implement the illegal fishing prediction system, such that we are able to detect the illegal fishing activities beforehand and then can take certain measures against it.

Currently, there are very few models or analytics that exist for detecting illegal fishing without the physical search and seizure by law enforcement human resources. Modelling both fishing behaviour and the illegal fishing enterprise will expose the data necessary to model and predict potential illegal fishing activity to focus law enforcement human resources on physically searching areas with a higher probability of detecting illegal activity. By using shipping vessels on board trackers, a methodology has started to be derived for describing typical fishing behaviour using data from legal fishers. Multiple institutions are collecting and refining this data as well as developing algorithms to detect illegal activities. Leveraging their work as a starting point and fusing these and other data sources an effective model can be developed to identify IUU fishing. Even with the available data trying to solve this problem globally is a very difficult task. It will be important to scope the project to a specific region, potentially targeting only certain fish populations to make the models more meaningful.

The main objectives are as follows:

- To avoid IUU fishing
- Provide efficient, Prediction system.
- Totally based on globally available data.
- Blacklist countries with negative behaviour
- Time saving.

7.4 TECHNICAL NOVELITY AND UTILITY

We used the regression model to identify the vessel's behaviour and determines if a vessel is normal fishing or not. This model adds some other new functionalities to the previous project and based on the given result law enforcement can know which vessel to investigate. we are focusing on the specific geographical area to identify illegal fishing.

According to the given data, every vessel has a unique id based on characteristics of vessels like we can identify whether it is a commercial good vessel, raw material goods vessel, etc. For every vessel, there will be a sensor attached to it so that we get data like the location of the vessel and the course of a vessel.

In this module user upload fishing dataset. Here we will get a pop up to upload the training data set. Data used in this model is collected from Global Fish Watch (GFW). The data was downloaded from GFW Sample Data. We have both numerical and categorical types of information in this dataset. Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviours or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. It loads at data from dataset.

Dataset is classified using build regression model. In order to predict the value of the dependent variable for vessels for whom some information concerning the explanatory variables is available. With the help of Knn-Neighbour model and uploading test data set we can predict whether it is a normal or illegal fishing this will be the final output.

By our model, we can predict if the vessel is fishing or not fishing. The vessel which is at high risk of doing illegal fishing will be shown in the result. So, whenever law enforcement finds such a vessel, they investigate them and stop another vessel by performing such illegal activity.

CHAPTER 8: CONCLUSION AND FUTURE SCOPE

8.1 INTRODUCTION

This chapter finally concludes the idea of our project. The system that is used will conserve the marine life and save environment which indirectly leads our global environment towards development. It is economical and easy to implement and replace the current system. The model which is used is totally based on regression which here works as a key factor for the data analysis. The system can be easily implemented on the global scale which needs this type of system as soon as possible. This section also walks you through the limitations of the system made by our team and the future advancement that could be done on the project.

8.2 CONCLUSION AND FUTURE GOALS

Our project concludes that It is our responsibility to protect our seafood and to keep many fishes alive. IUU fishing is one of the many to improve our fishing resources around the world and secure the health of our oceans. Now IUU's latest technology is used to monitor the vessel's location, their course, and records every detail of the vessel which enters the oceans. By the use of our project, we protect the unreported fishing and can find illegal fishing vessels. By mear of SAR satellite, we can continuously monitor the geographical location and can record every detail. to keep monitoring such activities even google has formed the GFW (Global fishing watch) utilizes all the activities to stop such illegal activities. So employing all this, we can reduce unreported, illegal fishing.

As the system detects the region where the illegal fishing is being done, the law enforcement team does not need to visit every major harbour for investigation and they can easily catch the illegal fishers within a less time using this system. Thereby it reduces their time of investigation and they do not need to waste them as in the manual process.

This is a promising new tool for detecting, predicting and deterring IUU fishing. Future research will focus on expanding the analysis to other geographic regions, and for other illegal activities, not just fishing, for example, narcotics trafficking and piracy. If the “good” vessels and fleets continually exhibit anomalous multiscale behaviour’s across regions and events, then it is possible to create near real-time and global indicators of IUU activity, broadly defined.

This new information will improve enforcement of maritime laws, and ultimately the sustainability of our seas. IUU fishing detection is not an actively attended topic yet it is a very deep and necessary topic to be studies with, our oceans is a main part of our ecosystem and the oceans are well maintained by its marine life. Hence, we need to take measures regarding this issue, fisherman doing IUU fishing can cause an instability in the ecosystem which will affect everyone of us, hence detection and stopping of these activities needs to be stopped.

8.3 LIMITATIONS

The model works on pre posted data and hence it predicts the IUU fishing activities based on that only, due to which it lacks the real time prediction feature. Also the data needs to be updated time to time for the accurate predictions. This model only predicts the IUU fishing hence to stop it we need the help of law enforcement or some kind of authorised official organisation to check on those activities and stop them from happening.

8.4 FUTURE GOALS

1. To improve the efficiency of the results coming, currently it is at 80% but we will try to improve this accuracy.

2. To automate the process using several API's and automation technique for the real time detection of the activities.

3. To create a portal for user interaction, so that they can check the model results on their will.

4. add more complexity to the project by taking more important factors into consideration which can lead to illegal and unreported fishing.

5. To provide a visual representation of the data analysis done by the prediction model.

REFERENCES

[1] M.M. Galotto-Tebar, A. Pomares-Padilla, I.A. Czerwinski, J.C. Gutiérrez-Estrada, "Is the vessel fishing? Discrimination of fishing activity with low-cost intelligent mobile devices through traditional and heuristic approaches", Expert Systems with Applications, pp.117091, 2022.

[2] V.K.G. Kalaiselvi, Pushgara Rani D., J. Ranjani, Vignesh Kumar S. M., Mohana Priya S., "Illegal Fishing Detection using Neural Network", 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), pp.1-4, 2022.

[3] Klawikowska Zuzanna, Ujazdowski Tomasz, Grochowski Michał, Piotrowski Robert, "How High-Tech Solutions Support the Fight Against IUU and Ghost Fishing: A Review of Innovative Approaches, Methods, and Trends", IEEE Access, vol.10, pp.112539-112554, 2022.

[4]Sunarmo, Achmad Affandi, Surya Sumpeno, "Clustering Spatial Temporal Distribution of Fishing Vessel Based lOn VMS Data Using K-Means", 2020 3rd International Conference on Information and Communications Technology (ICOIACT), pp.1-6, 2020.

[5] Kazushi Motomura, Tomoharu Nagao, "Fishing Activity Prediction from Satellite Boat Detection Data", 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp.2870-2875, 2020.

[6] Marza Ihsan Marzuki, Philippe Gaspar, René Garello, Vincent Kerbaol, Ronan Fablet, "Fishing Gear Identification From Vessel-Monitoring-System-Based Fishing Vessel Trajectories", IEEE Journal of Oceanic Engineering, vol.43, no.3, pp.689-699, 2018.

[7] https://globalfishingwatch.org

[8] https://iuufishingindex.net