EXECUTIVE SUMMARY

The report emphasises the critical role of implementing visualization techniques and classification algorithms in advancing fisheries management practices. It outlines how visualization techniques are instrumental in unravelling complex datasets concerning species diversity and population dynamics of aquatic ecosystems. By employing visualization in the environmental monitoring of fisheries, the report advocates for improved stakeholder comprehension, shedding light on fisheries' environmental trends and identifying areas or habitats in need of ecological interventions. Furthermore, the report highlights the role of classification algorithms in forecasting factors such as fish growth rates and stock abundance. The innovative predictive analytics approach presented in the report enables the formulation of proactive measures against fish stock depensation. Thereby, contributing not only to the sustainability of fish populations but also to maintaining the ecological balance of aquatic environments. A key insight from the report is the synergistic potential for combining visualization with classification algorithms. The convergence empowers fisheries and related stakeholders with actionable insights, fostering a culture of informed decision-making that underpins efficient fisheries' environmental and resource management. The report calls for the broader adoption of these analytical capabilities by identifying potential barriers and presenting a roadmap of recommendations to mitigate these challenges. In advocating these technologies, the report envisions a future where fisheries management is not only reactive but anticipatory, equipped with the tools necessary to safeguard aquatic biodiversity while simultaneously ensuring the sector's economic viability.

MOTIVATION

The global fisheries sector, supporting millions of livelihoods, is critically threatened with 80% of fisheries exploited or near collapse (TWC, 2024). This situation threatens biodiversity and food security and underscores the necessity for efficient fisheries management. Within this context, two challenges emerge, requiring a distinct analytical approach.

Despite the importance of freshwater fisheries, these ecosystems lack adequate attention compared to their marine counterparts. Freshwater fish are underrepresented in environmental monitoring initiatives, compared to other taxonomical groups (Radinger *et al.*, 2019). Freshwater fisheries are vulnerable to collapse, often owing to unidentified depensation or deficient compensatory recruitment of fish stocks, attributable to exploitation, predator-prey dynamics, or habitat degradation (Sass, Feiner and Shaw, 2021). Visualization emerges as a pivotal instrument to enable the identification of changes in species diversity and population dynamics, it substantially improves data comprehensibility, and integrity, uncovering patterns, trends and unexpected ecological shifts.

Moreover, the challenge of efficiently managing fisheries resources continues, notably due to industry practices favouring larger catch sizes (Ahti, Kuparinen and Uusi-Heikkilä, 2020). An alarming trend towards diminished fish sizes, with larger fish being supplanted by smaller ones, has been noted (UOY, 2023). This raises concerns for low-income populations where fisheries are vital to their nutrient security. Machine learning, particularly classification algorithms, provides a predictive solution to fish growth factors and influences stock abundance (Gladju, Kamalam J and A., 2022), ensuring the sustenance of both fishery-dependent communities and aquatic life.

LITERATURE REVIEW

The surge in environmental monitoring technologies has transformed how fisheries operate, leveraging tools such as GPS, network communications, digital cameras and image analysis software. This evolution generates dynamic, heterogeneous data that is both multi-dimensional and widely sourced (van Helmond *et al.*, 2019). Visualization has emerged as an indispensable tool to enhance data comprehension, stakeholder engagement, and identifying data collection gaps. It enables experts to detect inaccuracies in raw datasets and pinpoint areas in need of additional fisheries-related surveys or data collection efforts. Renowned entities, such as, the International Council for the Exploration of the Sea (ICES) have pioneered in deploying visualization to disseminate fisheries data, showcasing plots on catches, fish mortality, spawning stock biomass and stock status through various graphical representations (Anderson, Keppel and Edwards, 2020). As fisheries management increasingly involves stakeholder participation in governance, there is a greater emphasis on improving communication through visualization routes to convey scientific information. Visualization applications are instrumental across diverse environmental fishery domains, such as detecting patterns in extinction rates through pie charts, utilising geo- and spatial maps pinpointing regions requiring immediate conservation efforts, using line charts to forecast species that will be impacted by ecosystem trends, and using comparative charts into safety methods employed in catching specific fish types, and their repercussions on the environment (Volodymyr, 2021).

Parallelly, classification algorithms, a machine learning approach, have revolutionized fisheries management by offering predictive insights into resource challenges. These algorithms enable the rapid analysis of real-time data, integrate processing, and unveil patterns crucial for navigating the complexities of fisheries' data (Bradley *et al.*, 2019). Classification algorithms can be both supervised and unsupervised (Ahuja *et al.*, 2020), and their applications range from predicting fish disease through water quality analysis (Rana *et al.*, 2021), to tracking vessel movements for illegal fishing prevention, and managing stock depletion through market price analysis (Gladju, Kamalam J and A., 2022). Furthermore, classification algorithms predict fish species and their biological characteristics (Alsmadi and Almarashdeh, 2022), vital for sustainable resource management. It facilitates the estimation of fish size and weight through image segmentation and pattern recognition. It utilises decision trees for fish survival (Islam, Kashem and Uddin, 2021) and k-means algorithms for ecosystem analysis, showcasing the versatility of these methods in processing intricate datasets (Hastiestari and Syahidah, 2023). Classification improves decision-making and strategic planning, advancing the preservation and profitability of fisheries across crucial spatial and temporal dimensions.

CASE EXAMPLES

Visualization was employed to examine the environmental state of Lake Powell's fisheries. Two publicly available Kaggle datasets, one containing capture data of the fish in the lake and another cataloguing the contents within their stomachs (Appendix A), were merged. Using Rstudio for data pre-processing, these datasets were integrated, resulting in a dataset of 44 variables and 10,094 records (Figure 1, Appendix B). Date columns were standardized, outliers rectified, and categorical variables binarized to optimize dataset integrity from 2004-2018. Missing values were removed.

```
> print(variables)
 [1] "FISH_ID"          "EMPTY"        "ZOOPLANKTON"   "CRAYFISH"        "UNIDENTIFIED_FISH" "SHAD"            "THREADFIN"
 [8] "GIZZARD"          "CENTRARCHID"  "BLUEGILL"      "GREEN"           "CRAPPIE"           "BLACK_CRAPPIE"   "WHITE_CRAPPIE"
[15] "BLACK_BASS"       "LARGEMOUTH"   "SMALLMOUTH"    "STRIPER"         "WALLEYE"           "CHANNEL_CATFISH" "BLACK_BULLHEAD"
[22] "YELLOW_BULLHEAD"  "RED_SHINER"   "RAZORBACK"     "PIKEMINNOW"      "BONEYTAIL"         "HUMPBACK"        "CARP"
[29] "OTHER_FISH"       "ALGAE"        "AQUATIC_VEG"   "AQUATIC_INSECT"  "TRES_INSECT"       "Tres_Veg"        "DEBRIS"
[36] "OTHER"            "Flannelmouth" "Mosquito_Fish" "Snail"           "Muscle"            "DATE"            "TREND"
[43] "GEAR"             "PARASITE"     "Species"
```

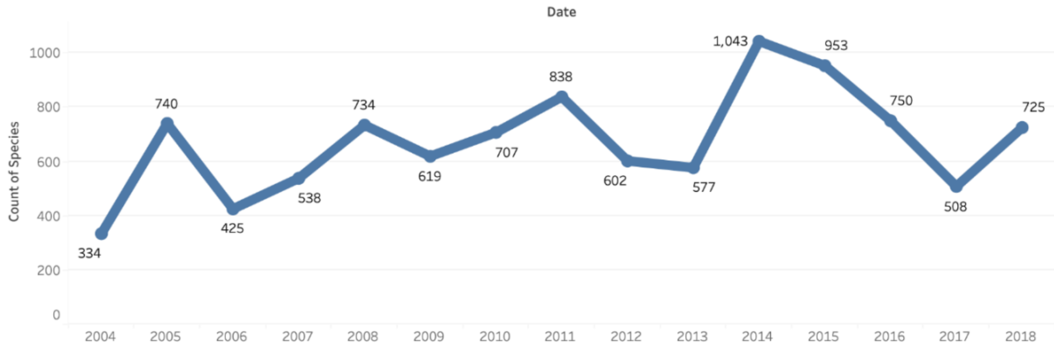
Figure 1: Dataset variables

Employing tableau to visualize fisheries data at Lake Powell from 2004-2018 facilitated a nuanced approach to environmental monitoring of this ecosystem. The analysis revealed a significant surge in fish captures in 2013, with Stripped Bass predominant (Figure 2). Examinations of stomach contents (Figure 3) yield insights into habitat conditions, showing trends such as the decline of instances of empty stomachs and fluctuating parasite infestations, with peaks in 2007, 2013 and 2015. Particularly, Stripped and Smallmouth Bass exhibited higher susceptibility to these parasitic encounters.

An increase in debris intake was observed in 2014, suggesting shifts in available dietary components or increased lake pollution. Concurrently, a discernible decrease in zooplankton from 2016-2018 raised ecological balance concerns and potential scarcity implications on the food web (Figure 4). Pronounced crayfish consumption, especially by Smallmouth Bass, underscored their abundance and dietary preferences, while the general dietary prevalence of ‘Shad and their relatives’ highlighted their ecosystem role (Figure 5). The findings illustrate the predominantly predatory nature among fish populations, preferring aquatic insects and algae, with a lesser reliance on molluscs (Figure 6). This example underscores the benefits of using sophisticated visualizations for understanding the intricacies and interdependencies of fisheries ecosystems.

Captured fish information

Number of captured fish in Lake Powell over time



Number of captured species in the lake over time

Species	2014	2015	2011	2016	2005	2008	2018	2010	2009	2012	2013	2007	2017	2006	2004
STB	462	474	240	327	319	178	312	308	120	232	342	315	184	273	140
SMB	239	229	320	172	206	296	186	171	200	150	68	75	141	50	95
WAE	269	166	151	135	124	132	160	153	215	136	122	48	116	56	80
LMB	64	80	76	44	40	90	20	39	68	39	12	43	41	21	8
CCF	9	4	13	50	44	26	38	32	6	31	28	41	23	15	9
YBH			8	11	3	5	7			13	5	6	1	3	1
BLC			22	5	1	2	2		10				1	7	1
GSF			6	5		2		1				1	1		
GZD						2				1		9			
CAT					3										
BGL			2			1									
WEA								2							
RZB								1							
BBH				1											

Figure 2: Captured fish data (see Appendix C for details).

Empty vs Parasite fish stomach

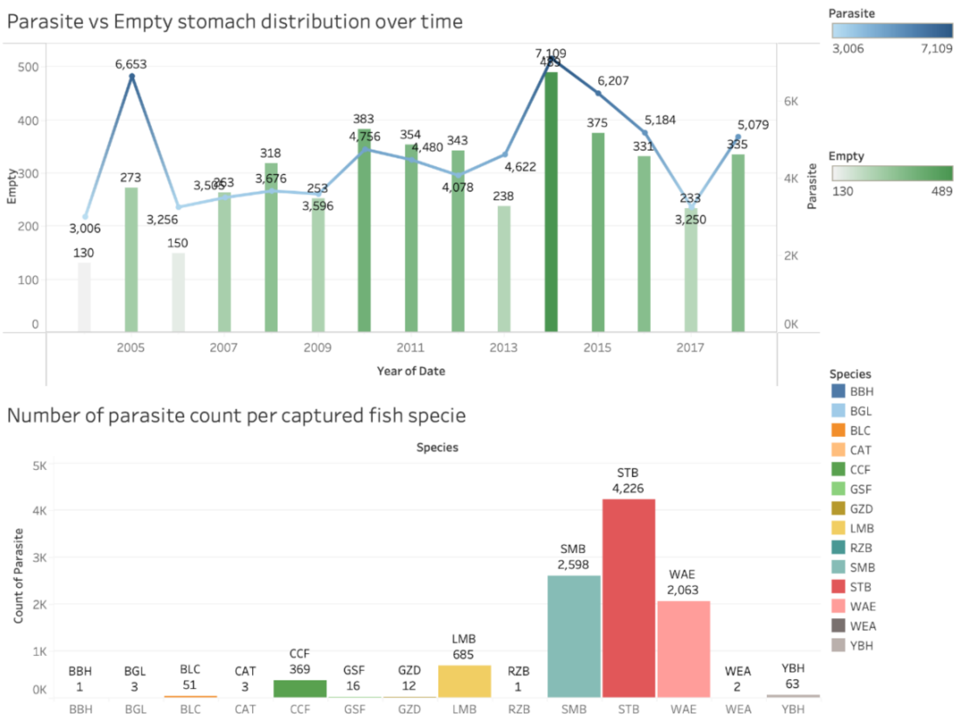


Figure 3: Empty vs Parasite contents

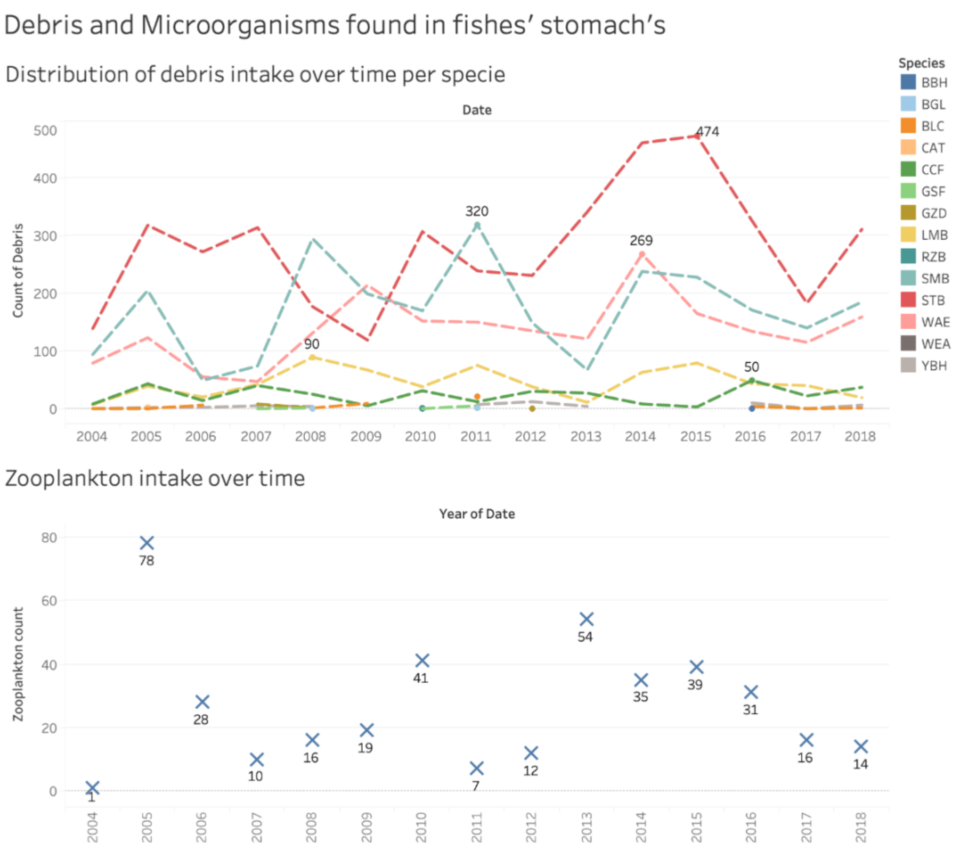


Figure 4: Debris and Microorganism Intake

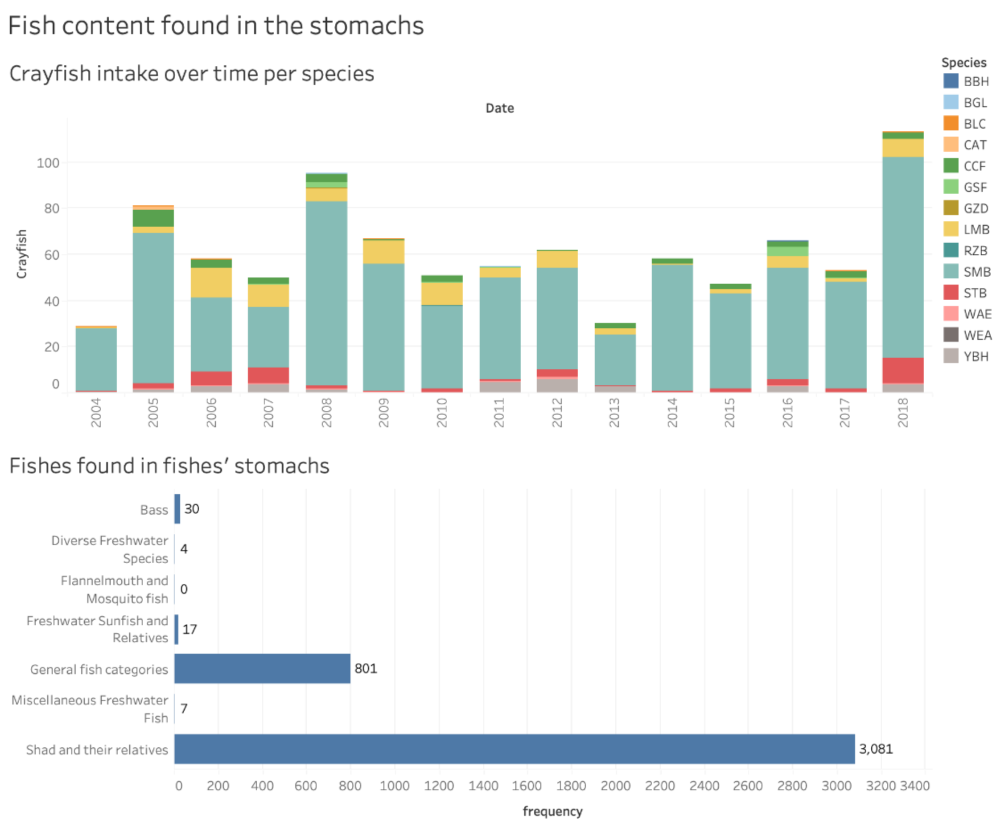
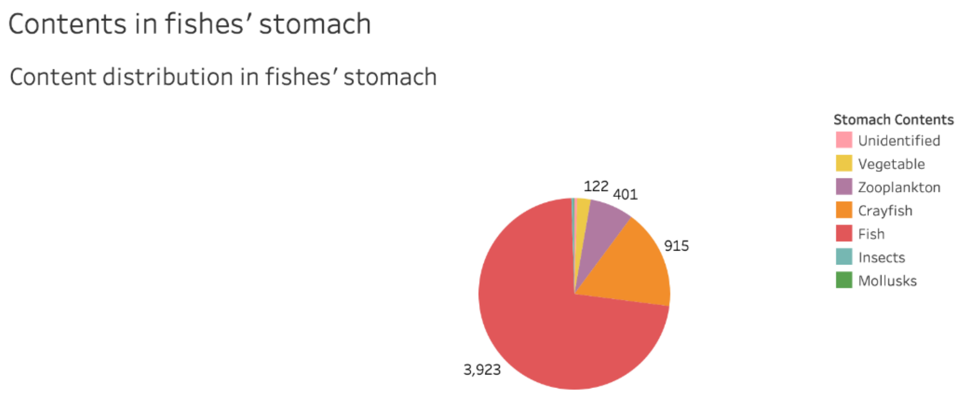


Figure 5: Stomach contents



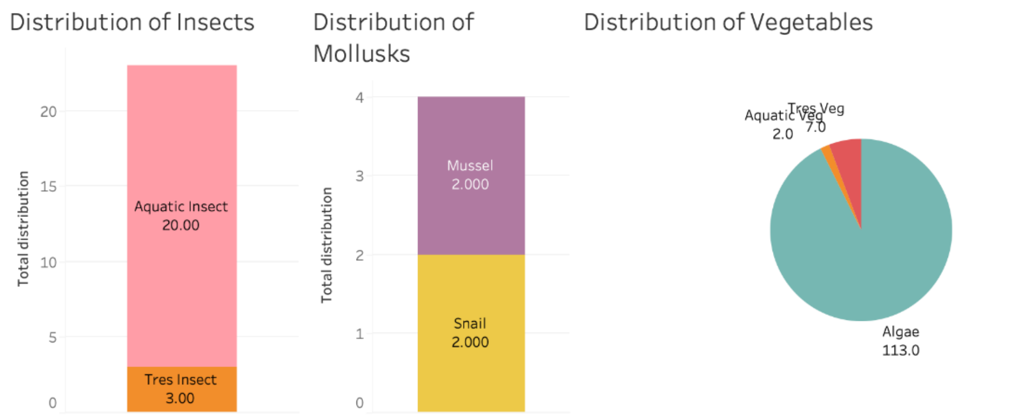
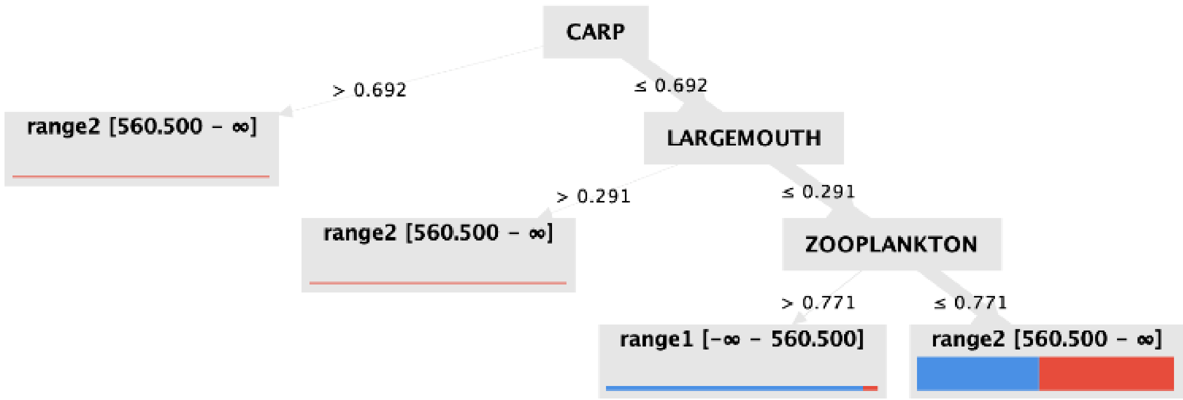
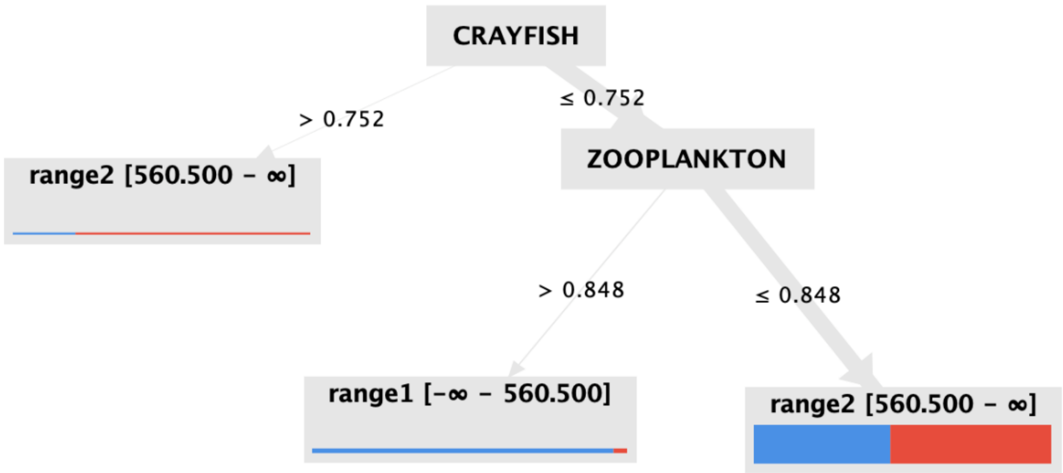


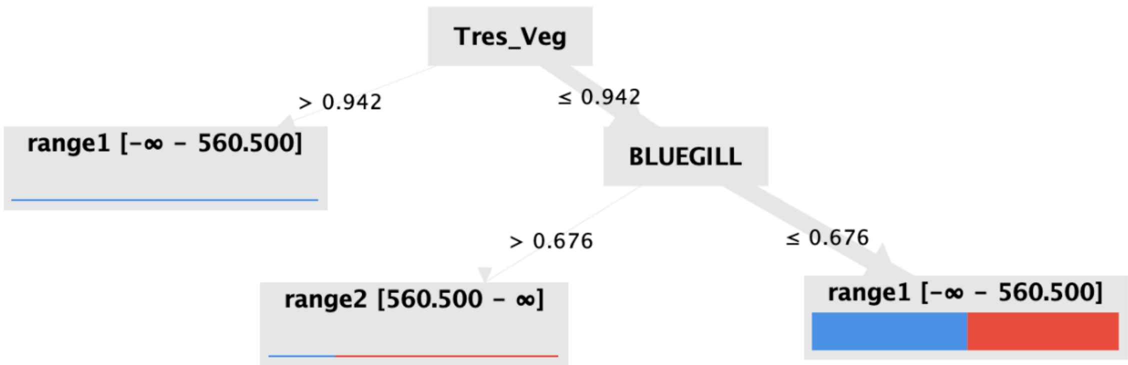
Figure 6: Distributions of stomach contents

Applying previous visualisation results, decision tree (DT) classification was employed to predict fish weight based on the stomach content of the predominant species, STB. The data pre-processing encompassed the exclusion of other species, incorporating a 'weight' variable, and eliminating uninformative and incomplete records in Rstudio (Appendix D and E). The dataset was further processed in RapidMiner by discretizing the 'weight' variable into two categories (72-560 grams and 560-2815 grams) to simplify analysis. The model, built on a random forest algorithm, utilized 'weight' as the predictor and employed cross-validation to partition the dataset for training and testing. Optimized with 100 trees and settings to manage overfitting, the algorithm's performance was enhanced, demonstrating effectiveness in handling complex datasets compared to traditional DTs (Appendix F).

The algorithm demonstrated substantial potential in predicting the weights of captured fish from their stomach contents, revealing ecological insights (Figure 7). It highlighted that larger fish consume more significant quantities and various prey types, including species influencing their growth like carp, largemouth, crayfish and bluegill. Conversely, lighter fish predominantly consume vegetation, including zooplankton and terrestrial plants. This discovery aids fisheries management by identifying dietary patterns linked to growth and predicting real-time food webs to intervene positively in ecosystems and increase stock abundance. The model achieved a precision of 94.77% for lighter fishes, signifying the model's reliability in making accurate predictions within this category, while a recall of 99.19% for heavier fish indicates a strong ability to identify instances in this group. With an overall accuracy of 56.91%, the model demonstrates competency above random chance, especially considering the high complexity of ecological data. Although the initial accuracy hints at potential enhancements, this foundation offers a significant step towards addressing and attempting to reverse the trend of fish reduction sizes, establishing a baseline for refinement with alternative classification algorithms.


CARP
> 0.692
≤ 0.692
range2 [560.500 – ∞]
LARGEMOUTH
> 0.291
≤ 0.291
range2 [560.500 – ∞]
ZOOPLANKTON
> 0.771
≤ 0.771
range1 [-∞ – 560.500]
range2 [560.500 – ∞]


CRAYFISH
> 0.752
≤ 0.752
range2 [560.500 – ∞]
ZOOPLANKTON
> 0.848
≤ 0.848
range1 [-∞ – 560.500]
range2 [560.500 – ∞]


Tres_Veg
> 0.942
≤ 0.942
range1 [-∞ – 560.500]
BLUEGILL
> 0.676
≤ 0.676
range2 [560.500 – ∞]
range1 [-∞ – 560.500]

accuracy: 56.91% +/- 1.10% (micro average: 56.91%)

	true range1 [-∞ - 560.500]	true range2 [560.500 - ∞]	class precision
pred. range1 [-∞ - 560.500]	308	17	94.77%
pred. range2 [560.500 - ∞]	1802	2094	53.75%
class recall	14.60%	99.19%	

Figure 7: Random Forest Variations and Performance Measures

BARRIERS

Technological challenges in transforming raw data into useful insights are due to the slow advancements of fisheries data capabilities in visualization and classification. Persistent challenges include the handling of datasets that are often incomplete, inaccurate or challenging to utilize, stemming from the complex nature of marine/freshwater ecosystems. This problem is exacerbated in recreational and subsistence fisheries, where data scarcity prevails. Significant data gaps remain concerning unassessed fish stocks, ecosystem impacts, sustainability metrics and conservation measures, along with inconsistencies in identifying aquatic species spatially and taxonomically (Blasco *et al.*, 2020). Moreover, computational tools are required to manage fluctuating environmental factors such as changes in light, visibility and ambient noise levels, to effectively monitor fish behaviour.

The implementation of sophisticated systems for data collection in fisheries also encounters practical barriers, including the high costs associated with technological acquisition, installation, and demand for skilled personnel (Bradley *et al.*, 2019). This issue is pronounced in small-scale, resource-constrained fisheries, where the financial burdens of procuring sensors, equipment and communication devices, coupled with the necessity for staff proficient in programming and analytical systems (Gladju, Kamalam J and A., 2022), deters adoption. Consequently, many fisheries revert to traditional practices to mitigate the risk of equipment malfunctions or technological failures.

Furthermore, concerns around data security and trust pose barriers for visualization and machine learning providers. Protecting sensitive information, such as fishing locations and resource databases, is crucial for commercial confidentiality (Gladju, Kamalam J and A., 2022). Yet, stringent security measures also foster resistance among fisheries wary of data misuse or increased governmental oversight. Additionally, the sector's unregulated nature and poor communication among stakeholders- such as policymakers, fisheries and other entities (Bradley *et al.*, 2019)- contribute to challenges like underreporting and illegal catches, complicating the adoption of these capabilities that could enhance transparency within the sector (Probst, 2020).

RECOMMENDATIONS

To harness the benefits of visualization and classification, fisheries must increase data richness by considering the long-term benefits of investing in cutting-edge tools, such as cameras, AI, sensors, IoT, autonomous drones, robots and augmented reality devices (Fujita, 2021). These technologies can enrich data quality, facilitating the collection of real-time data for efficient predictive analytics. The adoption of such technology demands collaboration through public-private partnerships, offering the required capital and expertise for the development and implementation of these data capabilities (Fujita, 2021). Additionally, encouraging a culture of data-sharing within the fisheries sector amplifies collective benefits and mitigates risks. Implementing information-sharing schemes that allow for swift gathering, analysis, and dissemination of data can help in accurately predicting and optimising visualization and classification models. Establishing trust is vital to ensuring that these programs do not lead to unfair competitive dynamics. Ensuring anonymity and providing assurance against penalties or restrictions is crucial for maintaining community collaboration (Calderwood *et al.*, 2023).

Adopting data capabilities in fisheries involves shifting to data-informed, technology-enabled operations, starting with leveraging existing data for immediate insights, such as existing catch volumes, GPS positions and fuel consumption records. Central to this transition is the need for comprehensive education, training and support, equipping fisheries' employees with the necessary skills, knowledge and processes, required to effectively utilize advanced analytics. Visualization and predictive analytics skills are emphasized, enabling the anticipation of future fishery trends and conditions (Lukambagire *et al.*, 2023). This holistic transformation, encompassing all functional areas, necessitates regular assessments of the initiatives' impact on costs, efficiency and sustainability which is vital to maintain momentum, promote awareness and secure further investment in analytics and data-driven practices. Moreover, fostering continuous dialogue and collaborative problem-solving among fisheries, data scientists, and stakeholders is crucial for aligning data standards and shared goals (Christiani *et al.*, 2019).

CONCLUSION

The limitations of my report stem primarily from the reliance on secondary resources, including market reports, government documents and academic studies, instead of first-hand, primary field resources. Furthermore, challenges inherent to fishery data collection, have restricted the availability of publicly sourced data, confining my case analysis to 2004-2018. Additionally, given the complexity of the dataset used, there is potential for algorithm-improved outcomes through the exploration of alternative classification methods or more advanced software tools.

The report underscores the importance of adopting advanced data analytics, such as classification and visualisation in fisheries management. The analysis stresses the pivotal role of advanced capabilities in revealing patterns, trends and ecological variations that are critical for informed decision-making, operational efficiency and greater sustainability of aquatic resources. Case studies presented illustrate innovative solutions using these capabilities to environmental monitoring and resource management challenges at freshwater fisheries, showcasing the suitability of these technological tools for the sector. The report advocates for the importance of technology-enabled, data-driven operations, supported by comprehensive education and collaborative efforts among all fishery-related stakeholders. As a result, this leads to a strategic approach towards improving performance and resource optimization, while addressing issues of overexploitation and endangerment of aquatic ecosystems, ensuring long-term viability and conservation of global fish stocks.

Appendix

Appendix A: Kaggle datasets

- AGE_GROW_TABLE.csv
- FISH_TABLE.csv
- LAB_COUNTS.csv
- STOMACH_TABLE.csv

Datasets chosen: Fish_table and Stomach_table, merged by the “FISH_ID” variable.

Source: https://www.kaggle.com/datasets/beamers/lake-powell-fisheries-data

Appendix B: Data pre-processing

```r
# merging 'fish table info' and 'stomach content' datasets
View(fish_table)
View(merged_data)
merged_data <- merge(stomach, fish_table[c("FISH_ID", "Species", "GEAR", "DATE", "PARASITE")], by = "FISH_ID", all.x = TRUE)

# chaning the date format
View(merged_data2)
merged_data2$DATE <- as.Date(merged_data2$DATE, format = "%m/%d/%Y")
merged_data2 <- merged_data2[order(merged_data2$DATE), ]

#capitalizing species variable
merged_data2$Species <- toupper(merged_data2$Species)


#converting FLASE/TRUE into binary
binary_variables <- c("EMPTY", "ZOOPLANKTON", "CRAYFISH", "UNIDENTIFIED_FISH", "SHAD", "THREADFIN",
                      +                           "GIZZARD", "CENTRARCHID", "BLUEGILL", "GREEN", "CRAPPIE", "BLACK_CRAPPIE",
                      +                           "WHITE_CRAPPIE", "BLACK_BASS", "LARGEMOUTH", "SMALLMOUTH", "STRIPER", "WALLEYE",
                      +                           "CHANNEL_CATFISH", "BLACK_BULLHEAD", "YELLOW_BULLHEAD", "RED_SHINER", "RAZORBACK",
                      +                           "PIKEMINNOW", "BONEYTAIL", "HUMPBACK", "CARP", "OTHER_FISH", "ALGAE", "AQUATIC_VEG",
                      +                           "AQUATIC_INSECT", "TRES_INSECT", "Tres_Veg", "DEBRIS", "OTHER", "Flannelmouth",
                      +                           "Mosquito_Fish", "Snail", "Muscle")

merged_data2[binary_variables] <- lapply(merged_data2[binary_variables], as.integer)

#outline of my new dataset
summary_stats <- summary(merged_data2)
str(merged_data)
print(dim(merged_data2))
variables <- names(merged_data2)
print(variables)

#checking for missing rows and correcting them
rows_with_missing <- which(apply(is.na(merged_data2), 1, any))
cols_with_missing <- which(apply(is.na(merged_data2), 2, any))
cat("Rows with missing values:", rows_with_missing, "\n")
cat("Columns with missing values:", cols_with_missing, "\n")
merged_data2 <- merged_data2[-rows_with_missing, ]
missing_values <- sum(is.na(merged_data2))
rows_with_missing <- sum(!complete.cases(merged_data2))
cat("Total missing values in the merged dataset:", missing_values, "\n")
```

Appendix C: Dataset Species

```
EL = Electrofishing
GN = Gill Netting
LMB = Large Mouth Bass
SMB = Small Mouth Bass
STB = Striped Bass
GSF = Green Sunfish
WAE = Walleye
CCF = Channel Catfish
CRP = Crappie
BGL = Bluegill
BLC = Blue Catfish
YBH = Yellow Bullhead
FWS = Flathead Catfish
RBT = Rainbow Trout
```

Appendix D: Dataset dimensions

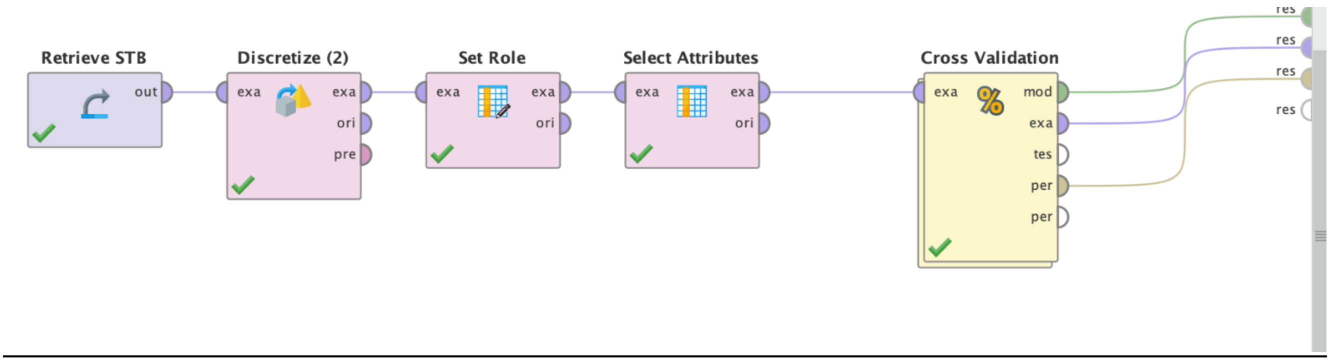
```
> dim(STB)
[1] 4221   41
```

```
> colnames(STB)
 [1] "FISH_ID"            "ZOOPLANKTON"      "CRAYFISH"
 [4] "UNIDENTIFIED_FISH" "SHAD"             "THREADFIN"
 [7] "GIZZARD"            "CENTRARCHID"      "BLUEGILL"
[10] "GREEN"              "CRAPPIE"          "BLACK_CRAPPIE"
[13] "WHITE_CRAPPIE"      "BLACK_BASS"       "LARGEMOUTH"
[16] "SMALLMOUTH"         "STRIPER"          "WALLEYE"
[19] "CHANNEL_CATFISH"    "BLACK_BULLHEAD"   "YELLOW_BULLHEAD"
[22] "RED_SHINER"         "RAZORBACK"        "PIKEMINNOW"
[25] "BONEYTAIL"          "HUMPBACK"         "CARP"
[28] "OTHER_FISH"         "ALGAE"            "AQUATIC_VEG"
[31] "AQUATIC_INSECT"     "TRES_INSECT"      "Tres_Veg"
[34] "DEBRIS"             "OTHER"            "Flannelmouth"
[37] "Mosquito_Fish"      "Snail"            "Muscle"
[40] "PARASITE"           "WT"
>
```

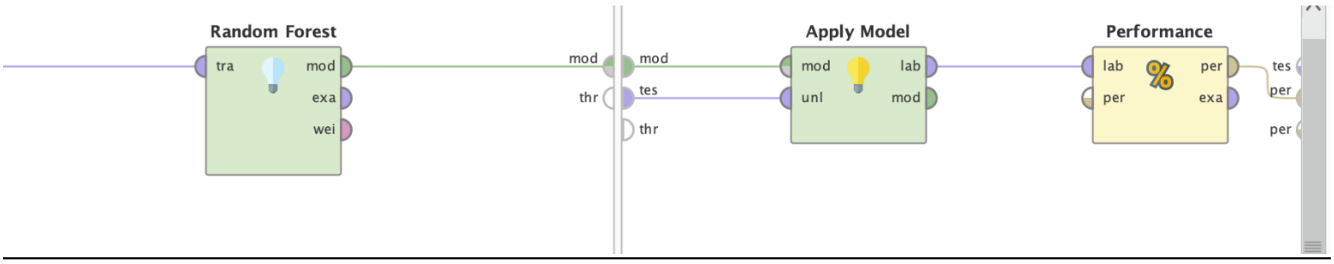
Appendix E: Data pre-processing for classification model

```
#CLASSIFICATION

# merge weight in fish table by FISH_ID in merged_data
result <- merge(merged_data, fish_table[c("FISH_ID", "WT")], by = "FISH_ID")

# identify missing values in WT row and delete them
result[, 46][result[, 46] == 0] <- NA
result <- result[!is.na(result[, 46]), ]

#delete TREND and GEAR columns
result <- result[, !(names(result) %in% c("TREND", "GEAR"))]
```

Appendix F: Decision tree process in RapidMiner



Retrieve STB
out
Discretize (2)
exa
exa
ori
pre
Set Role
exa
exa
ori
Select Attributes
exa
exa
ori
Cross Validation
exa
mod
exa
tes
per
per
res
res
res
res



Random Forest
tra
mod
exa
wei
mod
thr
mod
tes
thr
Apply Model
mod
lab
unl
mod
Performance
lab
per
per
exa
tes
per
per