|  |  |
| --- | --- |
|  | Abalone age prediction |
|  |  |
|  | - |

**Contents**

[**What?** 2](#_Toc170491435)

[**Why?** 2](#_Toc170491436)

[**Who?** 2](#_Toc170491437)

[**Goal** 2](#_Toc170491438)

[**Dataset** 3](#_Toc170491439)

[**Domain understanding** 3](#_Toc170491440)

[**Main question** 3](#_Toc170491441)

[**Side questions** 4](#_Toc170491442)

[**App prototype** 5](#_Toc170491443)

[**Reflection** 6](#_Toc170491444)

[**Sources** 7](#_Toc170491445)

# **What?**

The priority of choosing a dataset is where I could prove myself to the teachers while also finding something interesting or beneficial for myself to work on. During my search I found this dataset about blacklip abalones. One of the things you could do with it is predict the age of a blacklip abalones based on the rings on the inside of their shell. It sounds interesting, I never heard of the animal but quickly found out a few friends eat it in their household. Because I was so curious about it, and I wanted to know some fun facts for my friends I decided to dive deeper into the topic.

# **Why?**

The reason I choose this is because it sounds interesting to research more about an animal that I never heard about. While doing my research I realized the possibility what this tool could have. Retailers, researchers, fishers are just a few examples out of many who can use this. Being able to be the one that can make this tool sounds amazing. The problem, how I could predict the age with the help of machine learning sounds also interesting on its own. It boils down to the fact that curiosity and wanting to see the result of the model makes me eager about this topic.

# **Who?**

This can be a great use for many people. Fishers can set limits to make sure that younger and fertile abalones aren’t overfished with the help of this tool, making sure that the population isn’t in danger. Retailers can make use of to grade the abalones and set a price to it, since older abalones are worth more. It would help scientific studies by knowing the abalones age based on size so they can track an abalone behaviour, mortality rates, growth behaviour and so on. That way they can pinpoint when happens at what age. It would also help the abalones themselves. If we want to know a abalones age without a tool like this, the abalones need to be cut opened, stained and then through a microscope the rings on the abalone can be counted. This is a slow process which also kills the abalone. A tool like this would quicken the process and avoids killing the abalone. In short, many people from different branches can make good use of a tool like this. Even if the prediction is an estimation at best.

# **Goal**

The goal I have at least is that I can make a model that is able to make an estimation of a abalones age which comes close enough that one of the sectors I talked about can make use of it. However, what I really want and aim for is to make an age prediction where the error is at most in month, weeks. The question of that is if that would be possible with the dataset I got and if not, I need to know why. But it would be amazing to make an age prediction that performs this well. If that is happening, it would eliminate the method to slice open an abalone purely to record the age. More accurate predictions can also be used in many more situations. On top of that, sectors that could make good use of an estimation will benefit from this quality wise as well. For example, if a retailer knows if the abalone age is 12 or 13 years old it can get an idea on what to set the price too. However, if the person knows that the abalone is twelve but almost thirteen it would know what to set the price to.

# **Dataset**

<https://www.kaggle.com/datasets/rodolfomendes/abalone-dataset/data>

This dataset I collected by five researchers who did the following to gather the information which is this dataset:

* Review existing research.
* Gathering information at fishers
* Underwaters surveys.
* Tagging and monitoring abalones
* Studying their genetics

The research they performed started in 1982 and ended in 1984. Despite the age, abalones aging process did not change. The research they performed was a non-machine learning study and their findings are online to this day: <https://www.researchgate.net/profile/Warwick-Nash/publication/287546509_7he_Population_Biology_of_Abalone_Haliotis_species_in_Tasmania_I_Blacklip_Abalone_H_rubra_from_the_North_Coast_and_Islands_of_Bass_Strait/links/5d949460458515202b7bf592/7he-Population-Biology-of-Abalone-Haliotis-species-in-Tasmania-I-Blacklip-Abalone-H-rubra-from-the-North-Coast-and-Islands-of-Bass-Strait.pdf>

The dataset is donated in 11-30-1995 by Sam Waugh. However, the ownership of this dataset are the following parties:

* Marine resources division
* Marine research laboratories - Taroona
* Department of Primary Industry and Fisheries, Tasmania

The dataset got no missing values. It contains 4176 rows where each row represents an abalone. It got nine features and one target which is the feature “Rings”. Age is determined by adding a 1.5 on top of the counted rings.

Currently the dataset shows information about a abalones total weight, weight of separate things like gut, meat and shell. Length, height and diameter are there too which are the key features in the age prediction and the gender of the abalone is present of well. However not all abalones got a gender. Some are marked as infant because abalones gender cannot be identified when they are too young.

# **Domain understanding**

For domain understanding it is important to understand the topic and dataset. It will be required to do research, experiment with the dataset and come up with crucial questions. Hopefully I am able to find someone to interview but I need to see who seems the best to approach. Under “main question” and “side questions” you can find any noteworthy questions for this project.

# **Main question**

What do I need to make use of the regression dataset to accurately predict an abalones age?

# **Side questions**

* What is a blacklip abalone?
* How old do they get?
* How big can they get?
* How heavy can they get?
* Are there physical features to a blacklip abalone which is unique to them compared to the other variants?
* How can the dataset be used so that the result will be useful to at least one sector?
* Does an abalone stop growing?

**Domain understanding and research**

After doing my research I understand more about abalones and can approach the problem more effectively. Abalones are a mollusc who carry a shell. They are found in several places around the world, primarily coastal waters in in environments with a lot of rocks. Atlantic Ocean, Pacific Ocean and the Indian ocean are the ones where they are most known to be found.

The abalones are valued for their meat and shells. They got growth rings, just like tree. By counting them you can estimate their age, but it is almost impossible to not kill them to do this. Abalones do not stop throwing but at a certain age their growth progress slows down. Which is why measuring them is an ideal way of knowing the age. Their weight also grows in relation with the abalone.

The environmental condition, diet and health of an abalone can of course impact the weight and size. Male and female abalones got different growth rates which is something that needs to be kept in mind during the development of the model. When the population becomes too dense, abalones tend to fight over food which will impact the growth rates.

Now coming back to their size and weight the highest recorded values are as follows:

**Max age:** 50 years

**Max length:** 200 mm

**Max height:** 60 mm

**Max total weight:** 2300 grams

**Max shucked weight:** 1000 grams

**Max shell weight:** 800 grams

**Max viscera weight:** 800 grams

Now what is important to mention while answering these questions is that some of these max values also include exceptionally high but recorded size and weight. However, it is important to know it is possible and that it is not rare but rather uncommon.

The unique features of a blacklip abalone are the combination of the black lip margin and the shell characteristics. On top of that their preference for a specific habitat makes them stand out more compared to other abalone species.

Now to make sure that we make good use of the dataset in order to get good results. We need to examine the data. Check for patterns, certain values that stick out, check if it all is balanced, if not I might have to use things like SMOTE to balance it out. The hyperparameters need to be tuned. New features are required like circumference or BMI. Tests need to be performed with several models and ensemble methods.

Cross validation can be used to get a good performance on all classes. I would need to see if I can predict all classes. The max number of rings is twenty-nine. There is too little to go off with ages of twenty-five and above to actually make something out of it. There is a lot to check and try to get the best result. And it might not even be enough. On the dataset website it says that additional info like weather patterns, location and food availability might be required to solve this problem. All this information does not come with the dataset.

For the main question, what do I need to do to make a model that can accurately predict abalones age with this regression dataset. The answer sounds a bit too general but being critical Is what will help me. Let’s say a result came out of a model. I would need to ask, things like what ages it did well with, with what ages did not it too well, why is that what features did it use to come to this conclusion. With each question I actively need to get an answer. I need to understand why things succeeded and failed. I need to understand how the model thinks. I need to understand if it works for every class well or not.

If I keep finding problems and the cause, I would be able to improve the accuracy step by step. Another answer is that I need to make use of tools to make up for the imbalanced age and to deal with exceptional sized abalones. I think those two things will be the biggest obstacle.

# A screenshot of a computer Description automatically generated**App prototype**

Despite looking incredibly simply. This would be the idea for the program. An environment where an abalone requested information will be filled in and when the user is ready, he/she can click confirm to let the age predict. The reset button is purely there to clear out all the fields. The program will work best if all fields are filled in but if enough information is filled in and one is missing. For example, everything is filled but the length. The program will point this out and will be presented with the option to go back and edit the information or to proceed. Proceeding with an information missing will let the program calculate the average length based on its weight, diameter and height and then proceed to predict the age that way. However, doing so will make the prediction less accurate. But by doing so the user can still make use of the program if they somehow cannot fill everything in.

In an ideal setting the program would be hooked up to a device that can scale and measure the abalone which then will automatically fill in all the fields for you. This would save time and help with accuracy of the requested information.

Over here you can see what the message could look like if a field is empty.

A screenshot of a computer error

Description automatically generated

# **Reflection**

This will be a reflection on the work and research. A lot is learned and that is fair to say since I did not even know that these animals existed before I started on this project. The question now is of course: How is the model doing and can it be used in a professional way?

The model is doing fine. It gives out an average error of around 1.2 rings. Which is good enough for certain sectors like retailers and certain scientific studies where age phase Is what they want or good enough to name a few. So, the original goal I aimed for at least is fulfilled to a certain degree. There was little data on older abalones of a certain age. I tried to make use of SMOTE to balance them well with ages that we got more information off. But some ages got so little information that balancing it was impossible. Which is why it works to a certain extent, since it doesn’t support all ages.

However, this proves that it is possible to predict an abalone age and more information will certainly open the possibility to predict the age of older abalones and improve the accuracy on the abalones it is already predicant closely to. So at this point it can be use professionally to certain sectors.

# **Sources**

*Abalone Dataset*. (2018, July 19). Kaggle. [https://www.kaggle.c. datasets/rodolfomendes/abalone-dataset/data](https://www.kaggle.com/datasets/rodolfomendes/abalone-dataset/data)

Tasmanian Wild Fisheries Assessments. (2023, August 9). *Blacklip Abalone Biology - Tasmanian wild fisheries assessments*. Tasmanian Wild Fisheries Assessments -. <https://tasfisheriesresearch.org/blacklipab/biology/>

*Blacklip Abalone*. (n.d.). The Australian Museum. <https://australian.museum/learn/animals/molluscs/blacklip-abalone/>

*Blacklip Abalone*. (n.d.-b).

<https://www.dpi.nsw.gov.au/fishing/fish-species/species-list/blacklip-abalone>

*Blacklip Abalone | Fishing Tasmania*. (n.d.).

<https://fishing.tas.gov.au/species/abalone-blacklip>

Account, S. (2022, November 18). Blacklip Abalone. *Sydney Fish Market*.

<https://www.sydneyfishmarket.com.au/Home/Seafood/Species-Information/List/blacklip-abalone>

*UCI Machine Learning Repository*. (n.d.).

<https://archive.ics.uci.edu/dataset/1/abalone>

Wikipedia contributors. (2023, February 18). *Haliotis rubra*. Wikipedia. <https://en.wikipedia.org/wiki/Haliotis_rubra>

CandyAbalone. (2024, June 27). Blacklip Abalone | Exquisite Delicacies | Candy Abalone. *Wild Tasmanian Dried Abalone - Candy Abalone*. <https://www.candyabalone.com/blacklip-abalone/>

# 7he Population Biology of Abalone (\_Haliotis\_ species) in Tasmania. I. Blacklip Abalone (\_H. rubra\_) from the North Coast and Islands of Bass Strait

<https://www.researchgate.net/publication/287546509_7he_Population_Biology_of_Abalone_Haliotis_species_in_Tasmania_I_Blacklip_Abalone_H_rubra_from_the_North_Coast_and_Islands_of_Bass_Strait>