* Inspiration

As our team consists of different people that aren’t really familiar with or know each other it was a bit of a challenge to come up with an inspirational topic for everyone. So, we started out researching some available datasets that would be interesting and lined up with the topic of the hackathon. This theme being “Streamlining Solutions for Global Challenges.” We came up with multiple different options but quickly agreed and decided to pick and focus on a dataset for water quality classification. This because it fit well with the theme regarding global challenges, as access to clean and safe water is one of most basic the human needs. This is also shown as a part of the United Nations sustainable development goals namely goal 6: ‘’Ensure access to water and sanitation for all’’. Where worldwide challenges already exist today and that are only predicted to worsen with an even further increasing demand. The submission may not directly be linked to the scarcity or actual improvement of water quality as it is based on a fictional dataset, however we believe it could still help with the understanding and determination of water quality and safety with regards to consumption. This because the dataset consists of many different characteristics or elements in the water that affect the quality and thus the safety if their quantities are too high. The quantity threshold where water is considered to be dangerous and unsafe to drink differs between elements. These are also defined within the project.

* What it does

The project analyses a dataset of elements and attributes across ±8000 samples. These are preprocessed and ran through different models to determine if the sample is safe or not. This is done based on the measured amount of 20 elements across the sample. They all have their own specific threshold for what is considered dangerous. This means that if the level from one of the elements exceeds its threshold the entire sample would be considered unsafe and thus should not be consumed (without further treatment).

* How we built it

Perhaps not the most efficient way, but we decided to all individually approach the preprocessing process and then later on compared these to come to a final version. As there are different skill levels and familiarities regarding the machine learning aspect of the project some people had an easier time and were able to already expand upon this by utilizing different models as well. After the discussion there was some uncertainty on how to proceed and further improve the project with the remaining time. We eventually decided to expand by using some more additional models as well as to take good parts from the other codes to improve the most extensive version and also to start trying to increase the legibility by making the results more readable for the general public, by also implementing the results in graphs and images rather than just tables or a result line. Ultimately, we presented on a graph how the different models cope with and handle this problem and what their accuracy was.

* Challenges we ran into

Challenges were regarding the meetings and discussions where we were often initially not really sure what to pick, do or how to proceed, having a bit of trouble with actually making decisions. This brought down the efficiency as it took up quite some time. Some other challenges were the different skill levels and familiarity regarding the topic, implementation and machine learning in general. Some people worked on something similar before and some never did anything with it. While one of the goals is to also learn from the project and experience this made it more difficult for everyone to participate equally.

With the program itself we had some problems with different formats of data in our dataset, but we quickly resolved this issue. We also had some struggles with visualizing the data, but fortunately we came up with some ideas that we were satisfied with.

* Accomplishments that we're proud of

Despite some initial trouble regarding actual decision making and the different people in the team across different countries and time zones we still worked well together and reached multiple high accuracies across different models. Furthermore, we all learned something new and had a great time working on this machine learning problem.

* What we learned

This depends on and is different from person to person, but we learned more about classification in machine learning and teamwork regarding working together with new and unfamiliar people. We also learned how to use hyper parametrization, data preprocessing and other things that make our models work as well as they can.

* What's next

This is difficult to answer as the project has a fictional dataset and thus is not necessarily a reflection on a real-life problem or situation. But it could be nice if other people look at the project and learn from it, and eventually possibly implement it to help improve the situation and water quality somewhere in the world. In addition, thanks to this hackathon we can have some more background information and knowledge for other machine learning problems and in the future we could potentially solve more sophisticated issues.