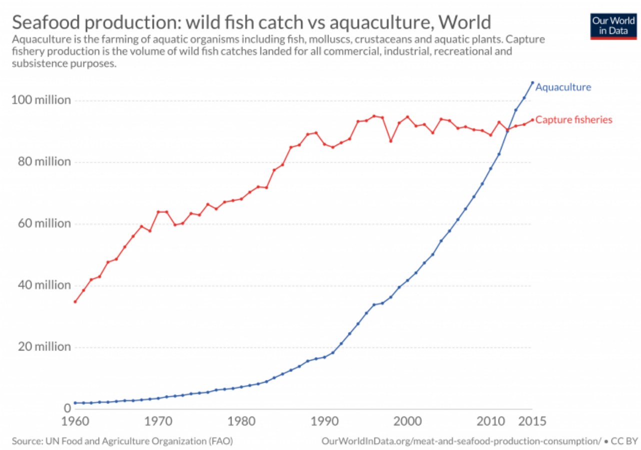
Using AI to make fish farming more sustainable at Aquabyte

As a machine learning engineer at Aquabyte, I led model development and data operations on our sea lice detection product, which helped fish farms report pest levels to the government.

Aquabyte's Underwater Camera



The growth of the aquaculture industry

A typical Norwegian Salmon Farm

Context

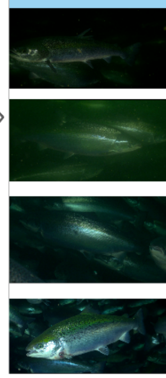
- As fisheries collapse globally, **Aquaculture has become the world's fastest growing food industry**, solving major challenges in food security and sustainability
- Because salmon farms have grown to house millions of fish, **populations of the pest sea lice have exploded**, risking farmers' crops and wild salmon populations
- **Every fish farm in Norway is legally required to report weekly sea lice counts to the Norwegian Food Safety Authority, the FDA of global aquaculture**
- **This process is currently manual**, tedious, expensive, and leads to undercounting due to small sample size.
- AI for precision farming has the potential to solve major food security and sustainability problems, so I joined Aquabyte, the world's leading Aquaculture AI startup. At Aquabyte, **I led development of computer vision and statistical models to automate sea lice counting**, I worked with the government to meet their standards for compliance reporting, and I collaborated with biologists, hardware engineers, and farmers to make precision farming a reality.

Building a sea lice counting pipeline: how can AI improve sustainability in the real world?

*Building on top of Aquabyte's hardware and infrastructure, **I built the entire end-to-end modeling pipeline for sea lice counting, visualized below.***

1. Fish Detector: Given full images, detect the individual fish in the image

2. Fish Filter: Given individual fish, determine which fish are visible enough to count sea lice.



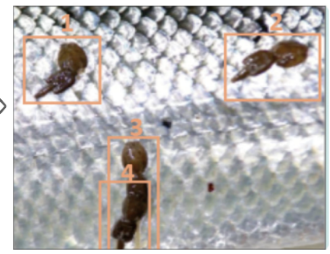
TOO DARK

POOR WATER QUALITY

OUT OF FOCUS

CLEAR ENOUGH TO SEE SEA LICE

3. Lice Detector: Given clear fish, detect sealice



4. Lice Demographic Model: Given individual lice counts, infer lice demographics breakdown

Chalimus Lice (Adolescent)

Observed

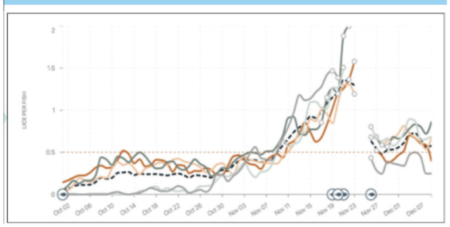
Inferred

Adult Lice

Government requires reporting lice counts broken down by stage of development, but **adolescent lice are too small to see using cameras**.

To solve this problem, I reproduced leading ecology research on sea lice demographic models to **infer** adolescent populations using **observed** adult data

5. Real-time lice data: Report average lice count to the farmer



From Research to Practice: After I built the end-to-end modeling pipeline for counting sealice, I needed to engage other stakeholders to put it into practice.

Addressing issues in the wild: New issues that hurt performance often emerged such as algal blooms, changes in fish behavior, or dirt on the camera. To support our field operations teams, I developed metrics and solutions to quickly identify, diagnose, and address these issues.

Integrating technology into government: In aquabyte's application to the government for automatic lice counting to replace manual counting, I wrote the technical methodology section documenting how our automatic sea lice counting product worked. This piece was a key component that led to the successful approval of our automatic method to be used for reporting to the government.

Mattilsynet

Norwegian Food Safety Authority

“This is a historic day”: First fish farmers celebrate dispensation meaning they can operate using only automatic lice counting

News by **editorial staff** - 24 November 2020

Using tech from Aquabyte, no more manual counting for Kvarøy Fiskeoppdrett and Seløy Sjøfarm. This will pave way for other salmon farmers to massively reduce fish handling.

Norway's Kvaroy, Seloy first salmon farmers exempt from manual sea lice counting

Other salmon farmers can now apply for automatic sea lice counting.

24 November 2020 13:03 GMT *UPDATED 24 November 2020 13:03 GMT*

By **Demi Korban**

Consigning manual lice counting to history?

REGULATIONS POLITICS SEA LICE HEALTH

F by The Fish Site
24 November 2020, at 11:49am

Aquaculture North America

MENU NEWS FEATURES PRODUCTS OPINION WOMEN EVENTS ENEWS MAGAZINE

Features> Disease Management Innovation

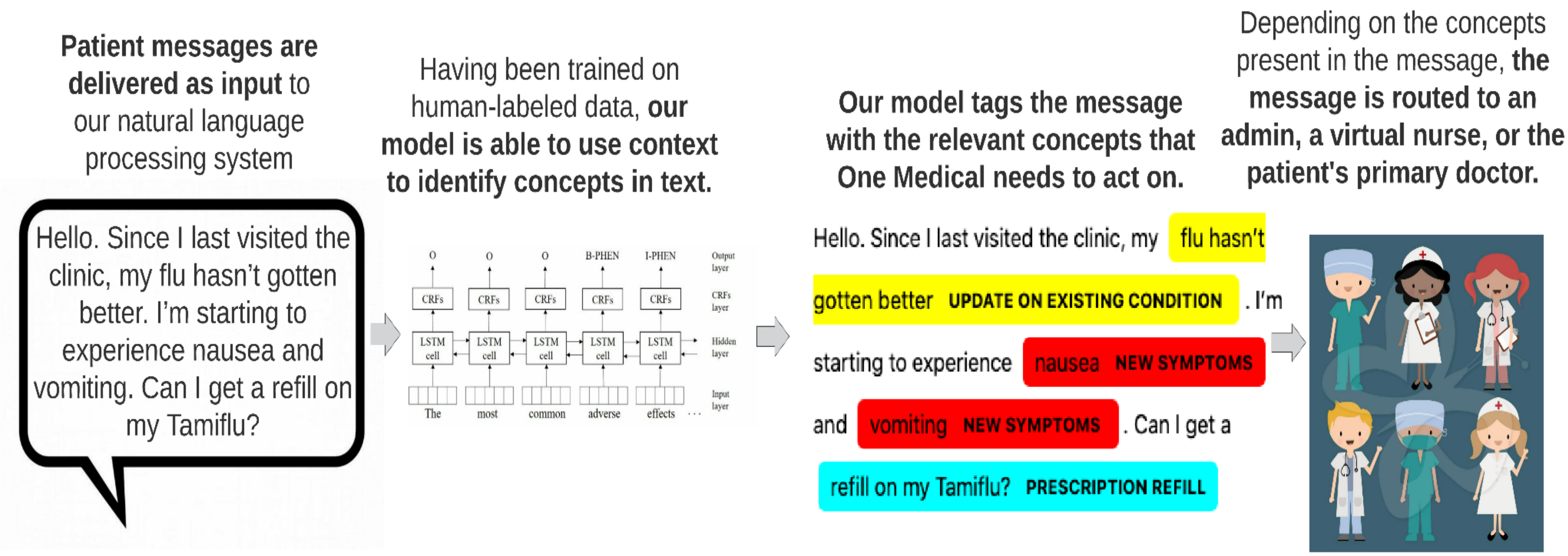
In Norway, counting sea lice manually may be on its way out

Impact

- **Policy change:** This system was the first and is still the only automated system formally approved by the Norwegian Food Safety Authority to replace manual reporting, dramatically elevating the role of AI in helping aquaculture become more sustainable.
- **Full Automation**: When I joined Aquabyte, they were serving one farm with manual work. Thanks to the automated pipeline I built, Aquabyte was able to scale to serving over 100 farms
- **Better pest data:** Sample sizes for farmer’s lice counts went from 10-20 fish a week to 50-100 fish a day, dramatically improving the quality of lice count numbers.
- **Improved transparency:** Better lice counts led to the discovery of widespread underreporting by fish farmers, forcing improvements in sustainable pest management.
- **Empowering farmers:** We saved significant work and time for the farmer, and empowered them with real-time data to better manage sea lice outbreaks.

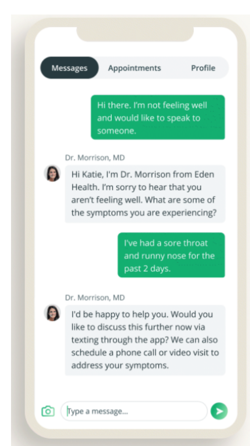
Using AI to make telemedicine more efficient at One Medical with Roam Analytics

As a machine learning engineer at Roam Analytics, I worked under Stanford Professor Christopher Potts to build natural language processing tools for healthcare companies. I led a project to build a telemedicine triage system for One Medical.



Context

- **As the health sector has digitized, we now have massive amounts of health text data.** I joined Roam Analytics, a health tech startup out of Stanford's NLP group, to seize on this opportunity to use text data to improve healthcare.
- ***One Medical* is a leading tech-enabled primary care provider in the US**. One of their key products is an app offering telemedicine services to complement in-person care
- **Due to the huge volume of messages, doctors spend 2 hours/day responding to telehealth messages**, and patients are often left without a response for days
- To save doctors time by reducing their message load, **I built a system for One Medical that used natural language processing to triage patient messages and decide whether a virtual medical team, doctor, or admin was needed to answer a given message.**



Building a telemedicine triage system required a human-centered approach to machine learning

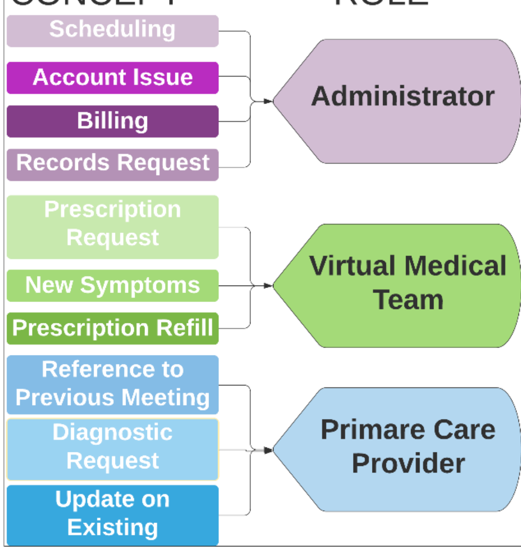
Scoping: Work with subject matter experts to define project requirements and a label schema

I broke the problem into 2 steps:

Find the relevant concepts: After working with nurses to review the data, I defined 23 concepts for the model to identify in patient messages

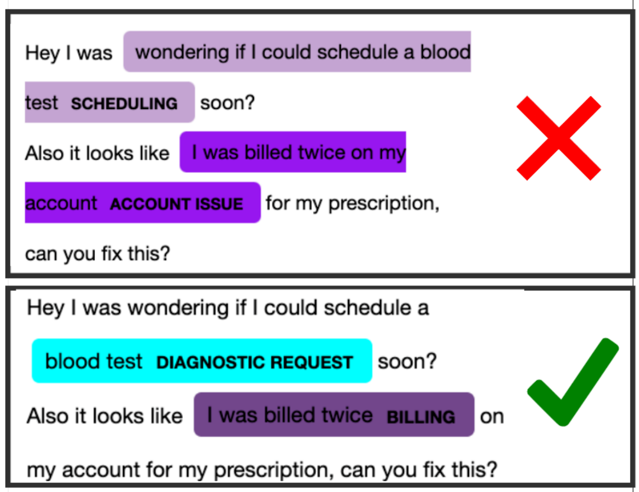
Match the concept to the role: In consultation with operational executives at One Medical, we matched each concept a patient might talk about to the role that should help the patient with it.

CONCEPT ROLE



Data Labeling: Now that we have our own label schema, we had to label the data ourselves.

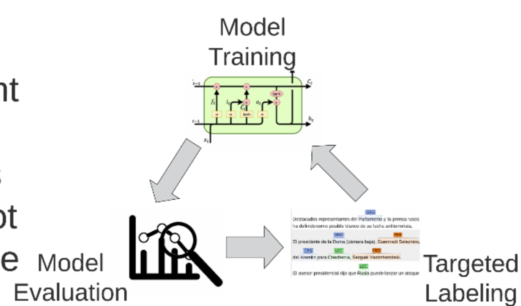
I wrote an annotation guide and trained nurses to consistently annotate our data. The correct annotation can sometimes be ambiguous, and training annotators to resolve these ambiguities consistently was essential to making sure our model performed well.



In the example, 2 annotators have labeled the same message differently. Though the message mentions schedule and account, the patient is requesting a blood test and help with billing, so the second labeling is right.

Modeling: I built a pipeline to train, evaluate, and improve the model

Given a labeled dataset, I implemented and trained a deep learning model (LSTM-CRF) to identify concepts in patient messages. Using a human-in-the-loop approach, I iteratively evaluated concepts where the model performed poorly and got our annotators to label more data for those concepts, culminating in a robust model.

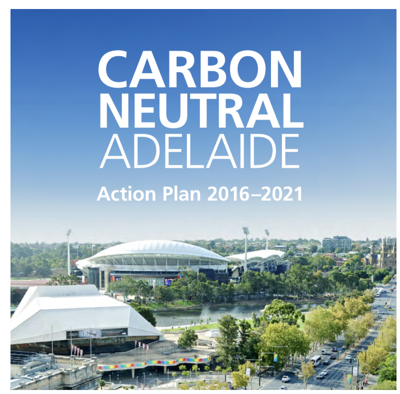


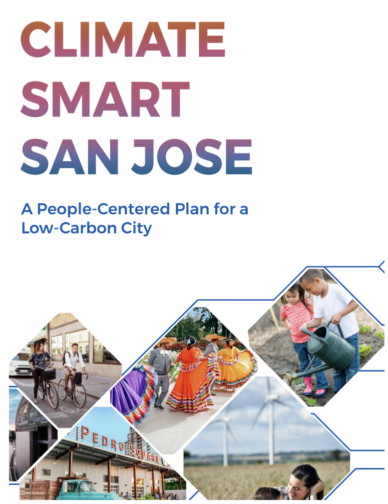
Impact

- 97% of messages were routed correctly in the real world.
- Our system was deployed to production, handling 30k messages/week.
- Doctors saw message load reductions of 56% saving 12000 hours a year, the equivalent of 6 full time doctors or almost 2 million dollars.
- One Medical awarded Roam a $200k contract as well as interest in other projects
- As One Medical grew, more and more concepts were split out into separate roles, such as billing and prescriptions. The flexibility of the concept role design meant the system could adapt to operational change.

Using AI to make climate policy analysis more scalable with Data-Driven Envirolab

Working with UNC professor Dr. Angel Hsu, I applied computational techniques to analyze cities' climate plans at scale. Our work culminated in the paper: *How are cities pledging net zero? A computational approach to analyzing subnational climate strategies*





Context

- As national governments delay significant climate action, **cities and states have taken the lead in setting net zero targets for emissions reductions.**
- **Cities' climate plans have low levels of standardization**, as they are usually released by cities as PDF reports with a wide diversity of formats, content, and structure.
- The large volume of cities' climate plans combined with their heterogenous format, structure, and content has made it **difficult for policy researchers to tractably compare large numbers of cities' climate plans**
- Passionate about applying my data skills to climate change, I approached Dr. Angel Hsu, then a Yale professor who ran the Data-Driven Envirolab, which applied data science techniques to climate policy research and emissions target tracking.
- **We believe natural language processing techniques grounded in social science have the promise of helping policy researchers scale their analyses.**

Grounding computational analyses in social science

Applying machine learning to research is wholly different from applying it to industry. Rather than automating some manual process, machine learning in research must empower researchers to better ask and answer scientific questions. As a result, the impact of this work must be described in terms of the research questions we aimed to answer.

Research Question 1: How are certain cities able to set more ambitious emissions reductions targets?

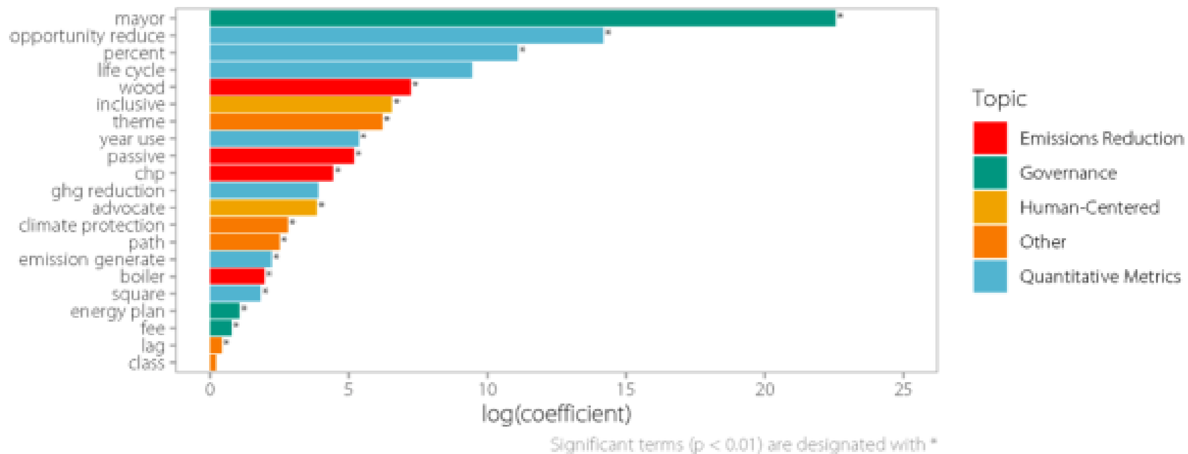


Figure 1. Terms predictive of economy-wide net-zero targets. P-values less than 0.01 indicate significant terms.

We found that "Ambitious" climate action plans - quantified, economy-wide net-zero emissions target of 80 percent or higher - have certain patterns associated with them.

- **Our study provided evidence that rigorous quantification of emissions inventories is key to successful climate plans, a finding that supports those advocating for better data and measurement in city government.** We found that discussion of metrics like emissions reductions and reference year emissions are associated with more ambitious emissions targets. While this conclusion may seem obvious, historically the connection between urban emissions inventories and climate plans has been less certain.
- **Our findings reinforced work from the previous literature that "active local politicians" and "a supportive local community" are key enabling factors of ambitious climate policies.** We found that governance-related language, such as the involvement of citizens in addition to top-level leadership from mayors, were also predictive of ambitious targets.Through manual inspection of the individual plans, we find that some of the ambitious cities' plans feature specific mention of mayoral support.

Research Question 2: What kinds of emissions reductions are cities prioritizing?

Climate policy researchers need to compare how emissions reduction strategies target different emissions sources. I applied text mining techniques to detect mentions of different emissions sources in climate plans (Top of Figure 4). Using this data, we constructed a detailed snapshot of how each city thinks about different emissions sources. For example, Figure 4 illustrates that while coastal New Bedford is heavily focused on water-related environmental issues, technocratic Munster emphasizes transportation and energy issues.

Figure 4. Topic word clouds for the cities of a) Munster and b) New Bedford. The size of each word corresponds to its tf-idf score or relative frequency of that term in that cities' climate plan. The color corresponds to the topic associated with each term.

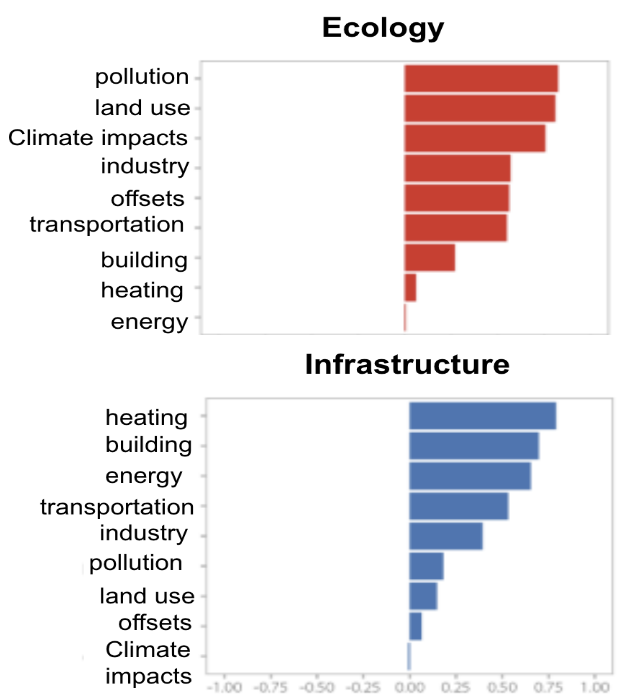


Figure 5. Two factors corresponding to Ecology and Infrastructure themes were determined from cities' climate action plans.

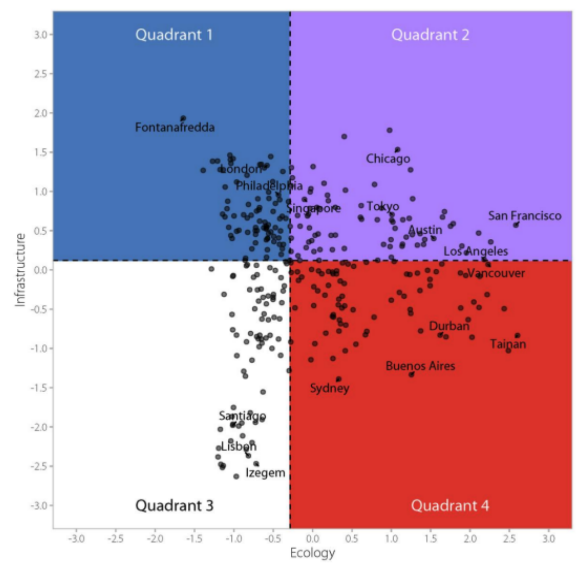


Figure 6. A plot of cities' factor scores according to two underlying latent factors: infrastructure and ecology.

We found that cities are trading off between infrastructure focused emissions reductions strategies and ecology related environmental solutions.
Using statistical techniques, we showed 2 factors, ecology and infrastructure (Figure 5), naturally emerged from the statistical structure of the data as a combination of individual emissions sources. This highlighted that cities are trading off ecology related environmental action with infrastructure environmental action, and many cities are overlooking broad categories of emissions reduction approaches (Figure 6).