

Capstone Project Proposal



Alex

Business Goals

Project Overview and Goal

What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you're labeling images, how will this help the business?

I'm proposing this as a project for a company which has developed a spirometry device. This device is used for measuring a person's lung volume and function. There are many such devices on the market, here is a random example of how they look like:



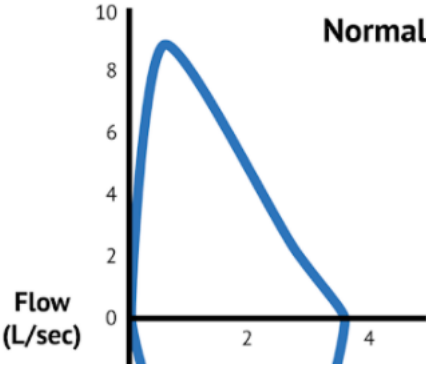
People with chronic lung diseases use to measure and monitor their respiratory function – it's used on a regular, sometimes daily, basis at home without a physician. What you basically do is exhale into the mouthpiece very strongly, until all air is out of your lung. Several indicators of your lung function will then be calculated by the device.

This company has developed such a device and added an online capability to it. The measurements it takes are uploaded to a server, and a physician can later inspect it on a web app. This product and the web app already exist, so we will build our ML service on top of it, as an additional feature.

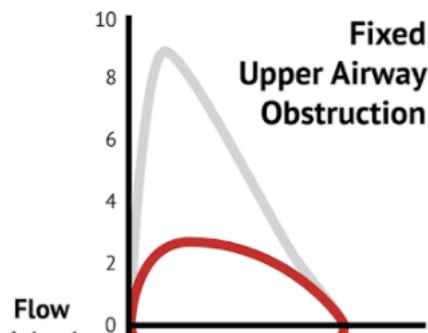
My proposition is to train a ML model to analyze the data of each measurement and decide right away if the patient has a respiratory condition or not.

Right now, this analysis is done by a physician, looking through all the data and trying to draw conclusions.

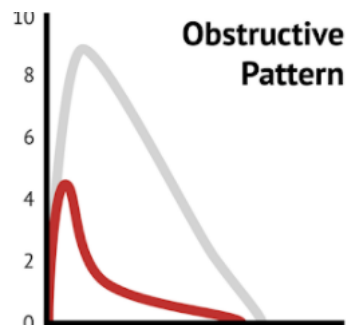
There are two benefits we can provide by applying ML to

	<p>this problem:</p> <ol style="list-style-type: none"> 1. Reduce the amount of time a physician needs to find relevant data to inspect (by filtering out irrelevant ones) 2. Identify the type of condition a patient might have
<p>Business Case</p> <p>Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.</p>	<p>This improvement will help drive sales by saving users valuable time.</p> <p>We have two groups of users: patients, and physicians. There's certainly a benefit for patients as they might receive better treatment – but the ones benefitting most are the physicians, as they'll save time and gain more confidence in their decisions.</p> <p>Very important to understand is that the physicians are the ones making the buying decision – they decide which device a patient uses (a patient could object, but in almost all cases they accept the physician's decision). Health insurances are then paying for the device (most people in the target geography have public health insurance), but they don't decide which one is used.</p> <p>So, physicians are the ones making the buying decision, and convincing them is key – with our ML-based analysis tool, we give them one more good reason to prescribe our device.</p>
<p>Application of ML/AI</p> <p>What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?</p>	<p>We will use ML to look into the dataset that was produced in each measurement and decide if the patient has a respiratory problem. The first condition we will look for in the MVP is an obstruction of the airway.</p> <p>This can be identified by analyzing one of the indicators produced by a measurement: the airflow per seconds, measured in l/s. For a healthy person, it looks like this:</p> 

For a person who has an obstruction in the airway, it will look like this:



Or this:



We can clearly see that there's not as much air flowing in these patterns. In many cases though, the differences are more subtle than shown here.

There are 4-5 such patterns that can be recognized from the airflow indicator that indicate an obstruction. The ML's objective will be to them, label them with the exact type of obstruction, and present them to a physician.

Success Metrics

Success Metrics

What business metrics will you

We want 90% of physicians already using the existing system use the ML feature on a daily basis (we can measure this usage via tracking on the web app). Further, we want to gain a 20% increase in sales

apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison.

through this product.

Data

Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	<p>The data will be acquired directly from the client I'm creating this project for. They have already gathered a large amount of these datasets from their active users, which we can re-use. The data will be free of charge for us, as we're already working with this client.</p> <p>We do not need to use any PII or otherwise sensitive data, only the raw measurements.</p> <p>Data will be produced on an ongoing basis by existing users of the system, so we can continue collecting more if needed. However, what we currently have should be more than sufficient.</p>
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	<p>For one, not everybody uses the capability of the inhalation device to send and analyze data. Some people will just use the device without allowing to upload the data (this feature needs to be enabled by the user, as it's very sensitive information that not everyone is willing to share. If data is not uploaded, the results of measurements can be read via Bluetooth and a mobile app).</p> <p>We can expect that younger people are more likely to use the connected features than older people, so there'll be a bias towards data of younger patients.</p> <p>Further, respiratory function depends on multiple other factors, such as height, weight, ethnicity, and physical fitness.</p> <p>There's a risk of people doing the measurement incorrectly. For the first try, one needs a physician nearby to assist, because the usage is not always intuitive. Afterwards it's easy to do alone. However, some people may still do it sloppily at least sometimes.</p>
Choice of Data Labels What labels did you decide to add to your data? And why did you decide on these labels versus any other option?	<p>We can use two different labels to start with:</p> <ul style="list-style-type: none">- Normal function- Airway obstruction <p>After that, we can fine-tune to identify different types of obstructions.</p> <p>And after that, we can pull in all other sorts of indicators other than the airflow. For example, common indicators are the total air volume exhaled, the volume in the first second and the first six seconds, and the volume for an</p>

	exhalation of one minute (extrapolated). All those can later be added for further diagnostics. In some cases, multiple factors need to be pulled together to come to a conclusion.
--	--

Model

Model Building How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?	We will train the model in-house, because it needs to fit specifically to our device and its medicine. We'll also host it in-house, as the resulting model may contain intellectual property which should not be shared right away. Our client already holds some patents though, so they might want to protect their model via additional patents and then host it on a cloud provider.
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	It's important that the model in this case does not miss any patients that may have a condition, so we'll optimize for recall – a few false positives are not as much a problem as are missed cases that have an issue. Because this is such a critical task, we should aim for a recall of at least 0.95

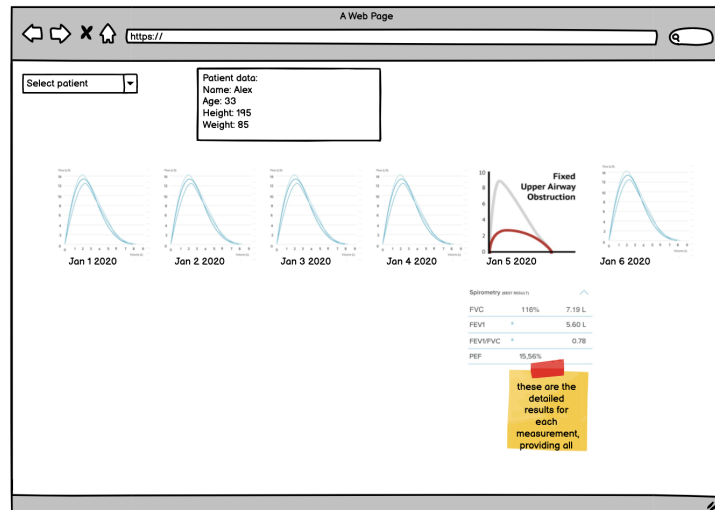
Minimum Viable Product (MVP)

Design What does your minimum viable	The minimum viable product is a web application in which a physician can select one of her patients. Upon selection, the measurement results are shown in a
--	---

product look like? Include sketches of your product.

row, ordered by the time they were taken (scrollable if there are many results). Those indicating an obstructive pattern are highlighted.

Upon clicking on a measurement, the physician can see all indicators that were taken, for closer inspection.



Use Cases

What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?

We have two personas:
Physicians and patients.

Physicians:

- Select one of multiple patients
- See all measurements in an overview
- Receive a recommendation from the model about which measurements indicate a respiratory condition
- Inspect a measurement in detail

Physicians will access the product via a web app, which they already use anyway (as we're building on top of an existing product).

Patients:

- Inspect their own data on a web app, without seeing the recommendation of the model (in order to not confuse them – a doctor should make the final call). This is just to provide them with a sense of transparency and control over what data is collected.

Patients need to take measurements regularly, which can be uploaded (that's happening automatically, given the device was configured to do so). This is already existing functionality, and not in scope of this project.

Roll-out How will this be adopted? What does the go-to-market plan look like?	We will roll out the MVP to doctors who already use the current application (which provides data only, without ML). Later on, we can convince more physicians to prescribe our device as of our improved functionality.

Post-MVP-Deployment

Designing for Longevity How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?	<p>In a next step after the MVP, we will add more lung measurement indicators to allow the model to identify other types of conditions.</p> <p>In terms of market adoption, we can cut out the ML part and apply it to data that was not generated by our device. We could offer an API for manufacturers of other devices, or of other medical software, to use. Another option would be to provide a web application solely for physicians, to which they can upload any measurements and get it analyzed. This would require some more effort but is certainly not impossible.</p> <p>The great advantage of such approaches is that we could gather a much more diverse range of data, which will improve our model greatly.</p>
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	<p>We have identified already that one of the main biases is on age – younger people are much more likely to use the connected features of this device. We can mitigate that by using the ML model in other products as well, which allows us to collect a broader range of data. For the MVP though, there's not so much we can do against it – other than collecting data from other age groups from other sources. This carries another risk though, as it might be hard for us to verify the quality of this data.</p> <p>There are other factors influencing lung performance, such height, weight, ethnicity, and overall fitness (though unclear how to measure that). We can monitor these biases by gathering such information together</p>



with the measurements taken (so a patient has a profile, into which they enter their personal attributes).