|  |  |
| --- | --- |
| 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? | Can we re-write this for spirometry?  A client I’m proposing this for is developing an inhalation device, which is used for treating chronic respiratory conditions.  Patients use it to inhale medicine, usually on a daily basis or even multiple times per day. It has an internet connection and collects and uploads data about the patient’s inhalation process.  A physician can then inspect the inhalation data in a web application. From the data, she can understand if a patient is using the device correctly.  Currently, physicians need to look through a large amount of inhalation sessions, look at the data, and draw conclusions. We can use ML to quickly identify if a patient didn’t inhale correctly.  There are two benefits we can provide by applying ML to this problem:   1. Reduce the amount of time a physician needs to find relevant data to inspect (by filtering out irrelevant ones) 2. Reduce the amount of datasets that slip the physician’s attention   As we will see on the next pages, using ML is well fit for solving this task – the problem is very narrowly specified, and the data can be labelled easily. |
| **Business Case**  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. | This improvement will help drive sales by saving the user valuable time.  We have two groups of users: patients, and physicians. There’s certainly a benefit for patients as they’ll receive better treatment – but the ones benefitting even more are physicians, as they’ll save valuable time.  The physicians are also 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 about it.  So physicians are the ones making the buying decision, and convincing them is key – we give them one more good reason to do so. |
| **Application of ML/AI**  What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve? | We will use ML to look into the inhalation data for each inhalation session, and identify those in which a patient didn’t inhale correctly.  There are three reasons why an inhalation is not done ‘correctly’:   1. Interrupted the inhalation 2. Inhalation is too short 3. Inhalation flow is regular (i.e. there are peaks in the inhalation)   Add drawing here |

**Success Metrics**

|  |  |
| --- | --- |
| **Success Metrics**  What business metrics will you 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. | Number of days when the inhalation was improperly done |

**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? | The data will be acquired directly from the client I’m creating this project for. Once they start to on-board real users, they’ll have enough data after a short while. I expect that after 90 days of usage, we’ll have enough data for a first MVP. We also have to consider, that we’ll have a larger amount of patients generating even more data. |
| **Data Source**  Consider the size and source of your data; what biases are built into the data and how might the data be improved? | For one, not everybody uses the capability of the inhalation device to send and analyze data. Some people will just use the device without ever connecting it. 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.  Further, there’s only one type of medicine usable with this device, so only one specific condition can be treated. So, our data will be highly biased towards patients with a certain condition. That should not be a problem though – we’ll train the model on this condition, and then only apply it on the same one. We just need to keep in mind that we can’t use the same model if we introduce another type of medicine for this inhaler. |
| **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? | I use three different labels, which all describe the ‘score’ or quality of a session of inhaling.  The scores are:   |  |  | | --- | --- | | < 50% | These are cases in which the inhalation has gone very wrong. If a patient scores these over a longer time, they seem to have severe issues with the usage of the device | | 50 – 80% | These are pretty good scores, but still not good enough in the long run. The patient is still doing it wrong. | | 80 – 100% | We can expect that if patients score these levels continuously, they’ll be using the device correctly and the medicine will be applied correctly. Should they still have problems, the treatment method itself is probably not helpful and needs to be changed. | |

**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 on our device and its medicine.  We’ll outsource the hosting of the infrastructure needed to a cloud provider, such as GCloud or AWS. |
| **Evaluating Results**  Which model performance metrics are appropriate to measure the success of your model? What level of performance is required? | Properly identified the periods in which a patient hasn’t inhaled correctly.  We should optimize for a high recall, minimizing false negatives. We want to miss as little problematic areas as possible – it’s worse to miss a day when they haven’t inhaled properly than to identify a day as problematic that was actually ok. |

**Minimum Viable Product (MVP)**

|  |  |
| --- | --- |
| **Design**  What does your minimum viable product look like? Include sketches of your product. |  |
| **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:**  They access the data via web app, where they can select their patients, and see their data on graphs. Problematic areas or time periods are highlighted and can be inspected in detail.  **Patient:**  A patient uses the inhaler on a regular basis. They too can see their progress on a web app – we could think of a further iteration in which patients see problematic areas themselves and even get recommendations. This is risky though – a patient might not be able to identify false positives or negatives the way a doctor does. So false recommendations from our model might have an impact on patients’ health. |
| **Roll-out**  How will this be adopted? What does the go-to-market plan look like? |  |

**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? |  |
| **Monitor Bias**  How do you plan to monitor or mitigate unwanted bias in your model? |  |