SEEKER

CLINICAL TRIAL PATIENT PROFILER

LIZL VAN JAARSVELD, 1612094

FABIEN CHIOTTI, 2381764

LEANDRA BRITS, 1608780

JARED MCCALLUM, 2304942

THE OPPORTUNITY 

In clinical trials, Doctors usually look at a list of patients and determine whether any of them will be suitable to participate in a trial by looking at certain biological markers. Doing this by hand can be extremely time consuming.

Who would benefit?

- Funders: this will decrease the cost in which the funders need to pay doctors to manually pick each suitable participant.
- Doctors: this will save doctors time, which they can now use to conduct the trials more efficiently
- Any individuals who will benefit from the outcome of the trial. Better selection for the trail may result in more accurate outcomes from the trial.

Why is this an opportunity?

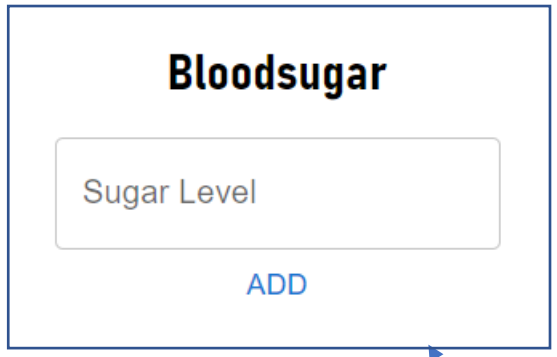
- Currently there are not many Well-known models which predicts whether a person would be suitable for a clinical trial or not. this creates an opportunity to expand on and improve this model.
- When testing a new product or medication, problems such as the cost of the trial and how to gather the participants play a major role. If the trial will cost too much then it often does not get done, or they find cheaper and less accurate alternatives to complete the trial.
- This will increase accuracy and productivity of the trial while decreasing cost and wasted time spent by doctors.

THE SOLUTION 

THE MODEL

We will use a Naïve Bayes model. This is a supervised learning problem because we can collect features with labels and then make a prediction (whether the person is a suitable candidate or not) based on the probabilities of the individual features. Our features will be the biological markers that doctors look for when they look for participants. Many of these biological markers will be continuous such as white blood counts, protein counts and antibody or antigen counts (see considerations). Many of these markers may also be discrete such as whether or not the patient has a certain disease. This is a classification problem, and we will choose our classes to be either 0 or 1, corresponding to whether the patient is suitable for this clinical trial (1) or not (0).

DATA COLLECTION



WHERE WILL WE COLLECT THE DATA AND APPROPRIATE LABELS?

Most hospitals already have a digital file system containing information about patients and their illnesses. We could use an app-scraper or an API to collect data from the already existing database. If their database is not conducive to our data gathering, we can include an add patient functionality to our product (see demo).

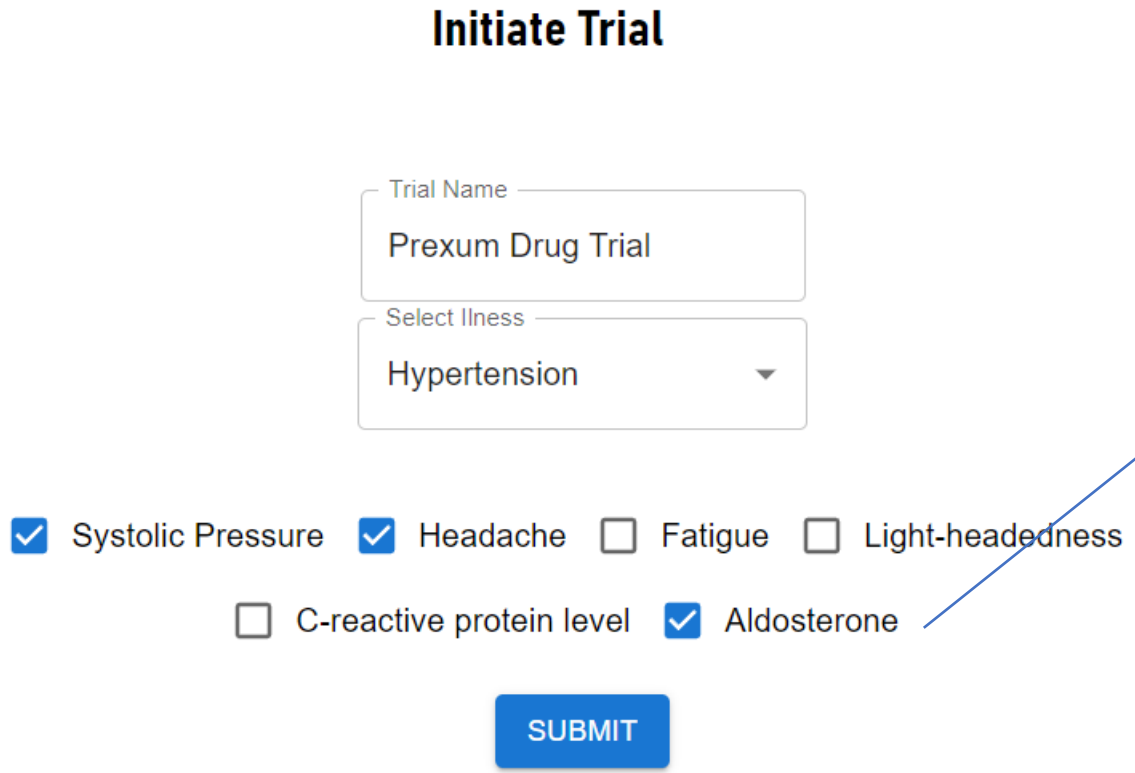
App-scrapers are easy to implement and cost-effective. Should an API be needed, or patients added manually , it would take no more time than doing it by hand, which would have been the default method. Furthermore, adding it by hand to the app once, will save you time if you had to check it again in the future.

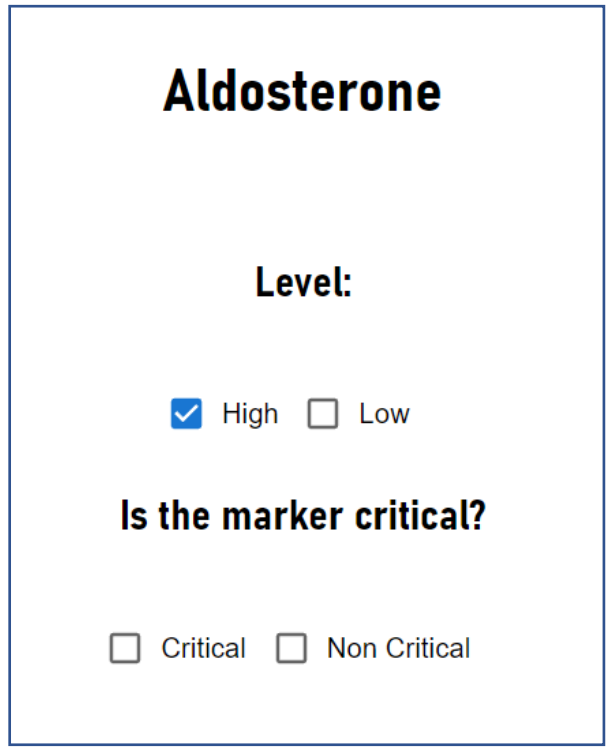
**Each illness in the select will have a pop-up pertaining to the information and biomarkers for the illness.

CONSIDERATIONS

- Data formatting: This will be the most time intensive part of the training process, however it should still be more effective than going through files manually each time.
- Feature choices: Although each disease has many features, there are usually key markers that are looked for to do the research required.
- Dataset size: Data sets may vary dramatically in different trials. We need to make sure that we have enough data in order to train the model correctly.
- Splitting the dataset: We need to ensure our data is meaningful when splitting the dataset into training, validation and test data.
- Encoding the continuous data: We need to convert any continuous data into discrete classes. This can be done by saying a value is either lower or higher than a reference value and depending if the preferred is lower or higher we will assign a 1 or 0 to it. This decision will be made by the doctor while choosing features (see demo).
- Smoothing: To overcome the 0-probability problem, we will incorporate Laplacian Smoothing in our model. It should be noted however that certain features immediately disqualify a patient from participating in a clinical trial as such, when choosing features, the doctor will be given the option to mark it as critical, we will add a column to represent whether a feature is critical or not, and if a feature is critical, it will not be smoothed.

FEATURE SELECTION





When choosing a feature with continuous values, a pop-up will be shown where the user can select a discretized form of the feature. For discrete features only a pop-up asking if the feature is critical will be shown.

OVER- AND UNDERFITTING

- Since Naïve Bayes is a probabilistic model, it as not as prone to over- and underfitting, by using priors, we moderate the fit. Laplacian smoothing is also a way to ensure the data is not underfitting.
- It is important to choose only representative features and to also ensure that our training set represents a multitude of cases. This would ensure that it does not lean towards certain features, although this is also mitigated by Laplacian Smoothing.
- An interesting thing to note that the Naïve Bayes model is relatively lenient towards smaller data sets. This works particularly well in clinical trials as certain illnesses are quite rare and might not have extensive data sets.

EVALUATION OF MODEL 

To prevent lawsuits on inaccurate predictions, the doctors will have to manually check the patients that are selected for the trials.

The goal of the app is to make suggestions of which patients to accept or not and is not sufficient to make the final decision.

This will however still save the time of going through the inappropriate participants. As the predictions are manually checked, the validity of the model is also checked.

DEPLOYMENT 

It will be a web-app which will have a remote database that stores previous trials with patient details used in those trials. The user will have the option to pick an existing model or create a new trial.

After the model is trained, it will draw participants not used in the training data from the database and output whether these participants are viable or not. The user will also be able to manually test patients not on the system, which leaves room for collaboration between hospitals.

HOW WILL OUR PRODUCT MAKE A PROFIT?

Each user(hospital or research organisation) will require a membership to access the data used in previous trials or to create new trials. Should the user wish to collaborate with other entities that are not subscribed, a commission fee will be charged. After the initial cost of creating the website, there would be little operational cost of running the website. Thus, profit will be quickly generated with each subscriber.

WHY SEEKER?

Seeker will speed up the clinical trial process significantly and will release pharmaceutical products quicker. Since there will always be clinical trials, there will always be an opportunity to do it more efficiently.