|  |  |
| --- | --- |
| 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? | The client I have in mind is a ride-hailing service – they pick up their customers within the city, at their home, and shuttle them to a nearby airport.  One of the challenges is planning their staff for the coming month – their drivers have flexible schedules, so the business wants to schedule them efficiently.  There’s an optimization problem here to be found (when to put which driver to make sure they can use their time most efficiently etc.) – which is not what we’re solving here.  What I want to solve is another question: how many trips will be necessary in the next month, so how many drivers will I need to schedule? |
| **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. | Customer success is certainly one factor – understaffing will lead to longer driving times (the driver has to pick up multiple people per trip, and if there are not enough drivers, each will have to take more passengers, with others possibly having to wait in the car).  A more important problem is cost efficiency. The drivers are paid by the hour, but they have to be scheduled for each month beforehand – and once they’re booked, they can’t be cancelled anymore.  So the more accurately they match their drivers to their actual demand (ideally on a daily basis), the lower their operational cost. |
| **Application of ML/AI**  What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve? | We will predict for upcoming months how many transfers and passengers will request a transfer. We will even predict the number of passengers and transfers on a single day. |

**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. | One possible metric can be the amount of passengers per transfer (for the vehicles they use, the maximum – and ideal - amount of passengers per vehicle is 8).  A constraint for this metric: The total trip duration for each passenger must not be longer than 90 minutes (where it already takes 45 to get to the airport).  One problem with this: if there are enough drivers but their routes are planned poorly, their efficiency will also suffer – and our success metric drop. So we need a better metric. |

**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? | We will use historical data of that company – they’ve been collecting all their bookings for 15 years already, and this data is still accessible.  There’s lots of personally identifying data in this dataset – but handling this is not an issue. We’re already working with it and have the necessary documentation and processes in place (the EU GDPR applies to that company).  The company is constantly updating (with each new booking), so we can continuously feed back fresh data into our model. There are more than 500k bookings available already. |
| **Data Source**  Consider the size and source of your data; what biases are built into the data and how might the data be improved? | tbd |
| **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? | We’ll label each day with the amount of passengers to be expected. |

**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? | Not sure if an outside platform will have such a model in place – we could do the labeling outside though.  We’ll then build the model in-house |
| **Evaluating Results**  Which model performance metrics are appropriate to measure the success of your model? What level of performance is required? |  |

**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? |  |
| **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? |  |