|  |  |
| --- | --- |
| 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 targeted with this proposal is a car sharing company. Car sharing companies have been generating loads of data during the last 10 years (whereas a majority of providers has only entered the market in the past five years).  A major hurdle for all providers is profitability. Operational costs are high, and vehicles must be loaded for as much as possible in order to break even. Clients on the other hand wish to have a vehicle as close as possible to where they need it.  Having a vehicle at the location where it is most needed as often as possible is thereby key to increase revenue as much as possible. |
| **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. | A major obstacle in adoption of car sharing models is availability of vehicles nearby – the farther a vehicle is away from a potential user’s location, the more likely it is for the client to choose another means of transport (or another car sharing provider that happens to be closer – this issue being amplified by the emergence of multimodal mobility applications).  Hence, the more exactly vehicles can be positioned at the location where they’re needed, the more revenue they can generate. |
| **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 the demand of car sharing vehicles in specific areas of a city, given historical data. |

**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. | We should aim to maximize precision, therefore reduce false positives. False positives may incur cost in case a false positive prediction will motivate a fleet manager to move a vehicle. False negatives (i.e., ‘no vehicle needed at location x’) might mean the location of the vehicle is not improved, but also no effort it lost by replacing it.  Therefore, maximizing precision will mean to only move a vehicle if there’s a high probability it will really be needed at it’s new location. |

**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 the client – they’ve been gathering this data for a long time, so there should be a lot of it available.  Wrangling the data will be a large part of the work.  There will be an ongoing influx of fresh data, which is a great way to measure the impact of the model, and to further improve it. |
| **Data Source**  Consider the size and source of your data; what biases are built into the data and how might the data be improved? | The dataset might be pretty large, depending on how long the client has been operating yet.  Biases:   * The business area of many car sharing providers is changing regularly * Adoption of car sharing has just taken place in the recent years, so very early data might be distorted * Mergers such as DriveNow and Car2Go might lead to further distortion * Certain events that occur irregularly will have a large impact. For example, Oktoberfest in Munich * Operational issues with the public transport will also have an impact on the demand that will be very hard to predict |
| **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 can start with a simple model:   * Split the business area into cells (we’ll need to find the right size of cells) * Assign each cell a value of 0 or 1 for a fixed period of time, e.g. every hour of a day or every day * 0 and 1 just means if a vehicle is needed or not * Later, we can count the amount of vehicles needed |

**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? | Model training will need to be done in-house – it’s probably also not a problem that can be solved simply by labeling.  It is likely that clients will be very secretive about their data – they won’t allow us to capture it. They’ll likely require us to train a model on their data, and have it run only on their own systems. |
| **Evaluating Results**  Which model performance metrics are appropriate to measure the success of your model? What level of performance is required? | We should aim for a precision of 90% or more. False positives are not desired. |

**Minimum Viable Product (MVP)**

|  |  |
| --- | --- |
| **Design**  What does your minimum viable product look like? Include sketches of your product. | The MVP shows a map with the business area, split by cells indicating demand for a fixed time period.  The user can move through the time periods of a day, or look at larger periods (e.g. 6h, 12h, 24h demand)  It is possible to zoom in and out of the map, and show dense areas as clusters. |
| **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? | The persona to use this system is a fleet manager. She’s responsible to ensure smooth operation of the fleet and make sure that clients can use the vehicles as smoothly as possible.  Further their job is to distribute vehicles (usually happens after a vehicle has been serviced, or if a vehicle has been place outside the business area). |
| **Roll-out**  How will this be adopted? What does the go-to-market plan look like? | We will need to find a test client first that is willing to let us use their data to train a model.  Once a model is trained, we might be able to use some of it on other clients, but mostly we’ll have to adjust it to their needs. Especially, we’ll have to groom the data of each client separately. |

**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? | New data is generated on a daily basis and can be used to improve a model.  We can improve the UI of the application and show a heatmap of some sort predicting the amount of vehicles needed at a specific place.  There’s another dimension which can be explored as well: Just because a vehicle is needed at a certain location, it doesn’t mean that the resulting trip is valuable. Longer trips are desired to shorter ones – we can take this factor into account and make a prediction of not only where vehicles are being needed, but how lucrative the resulting trip might be. |
| **Monitor Bias**  How do you plan to monitor or mitigate unwanted bias in your model? | Compare predictions of the model to actual demand |