Capstone Project Proposal

Md. Abdullah Al Mamun

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?	An Industry Problem: In Bangladesh, it's getting difficult to verify vehicle registration due to over vehicle in a city like Dhaka (Capital of Bangladesh). Currently, the system to verify vehicle registration, have in Bangladesh is very time consuming which is another reason of Traffic Jam in most populated city. This is not only a problem for citizens but also for the administration. The business solution is to automatically detect the vehicle registration number and verify the owner by making the detection part with computer vision. [N.B. In Bangladesh all vehicle registration plates are written in Bengali Language]
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 product will increase the citizen satisfaction as well as the authorities. Because, for the citizen they don't have to wait for long time and for the authorities they don't have to verify manually by checking some hardcopy papers. It will also help to decrease the traffic jam. It will increase the revenue of country and also decrease the corruption of government officers (in Bangladesh).
Application of ML/AI What precise task will you use ML/AI to accomplish? What	This might be done by using Computer Vision, in our case we will train our learning algorithm on Vehicle Registration Plates images to get the number which is written in Bengali language. Then we will use a web app

business outcome or objective will you achieve?	to verify that number from the record of government vehicle registration database. It will decrease significant amount of time to verify a vehicle.

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.	The success metric will be measured by monitoring the decrease in time to verify the vehicle registration and customer satisfaction. The customers include citizens and the authorities. The success metrics can also be measured by the rating given by a verifier officer.

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	At the beginning, we are going to take 10k images of vehicle registration plates and label them with third party. We will monitor and improve image labeling by using Appen. Initially we will only try to read the numbers of the vehicle not other texts in the plates. There is personal identifying information or data sensitivity issues. Every registration number is someone personal property. So, we have maintained privacy and ask permission to take photo of their vehicle registration

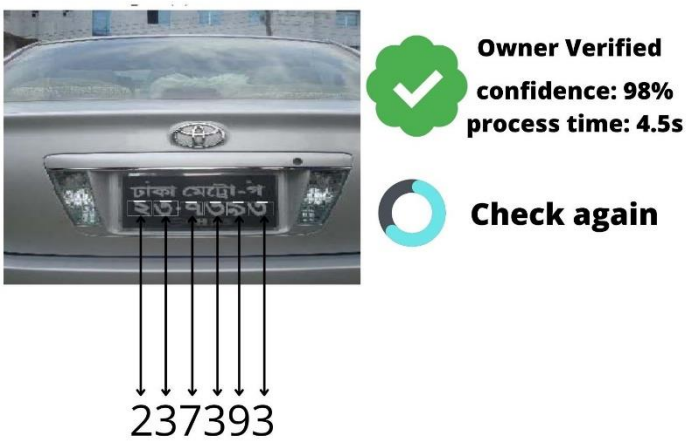
available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	plates and also need government permission to that work.
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	How the data will be collected? First, we will search for publicly available dataset. To be honest I didn't find any public dataset for vehicle registration plates which is written in Bengali Language. That's Why we need to hire photographers to take photos. Though it is costly but this is necessary. Yes, there is some issue with biases. We must collect photos which contains all 9 numbers (0-9) in a similar manner. Otherwise, it will be failed to detect the correct number. That's how we can improve our data.
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?	Initially, we need only 9 numerical character labeling which is 0-9 in Bengali. We will also recognize 50 characters of Bengali Language in the future for getting more information about the vehicle. The very simple solution will require a human to review the photo against the labeling.

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?	As I mentioned above there is a personal issue with the data that we are working on, so we need to keep our model private. There is another issue with computing power and memory. So, there is good a way to start with a small cost is to choose a cloud vendor like AWS to train our model. We will also keep inference or prediction system on the cloud. This will definitely help us to launch a minimal viable product. We will train the dataset on various higher tiers models such as OpenCV to get the best results. This process is iterative and with more that's, we will keep updating our model with time.
Evaluating Results Which model performance	**Character Error Rate**: - It is computed as the Levenshtein distance which is the sum of the character substitutions (Sc), insertions (Ic) and deletions (Dc) that

metrics are appropriate to measure the success of your model? What level of performance is required?	are needed to transform one string into the other, divided by the total number of characters in the ground truth (Nc) $$CER = \frac{S_c + I_c + D_c}{N_c}$$ The level of performance is required for this CER is at most 4% to 5%.

Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	 User interface: The initial interface is build on by react js and deployed on cloud. React JS will communicate our model through API. It will provide service to take instant photo and automatically submit then get the output.

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 interface will be very simple. Authorities will use it to verify with less time. They will take instant photo by this app and it will send automatically to the API and will get some output as a response. It doesn't need any user manual or tutorial to learn. Only selected user can access this app due to privacy issue.
Roll-out How will this be adopted? What does the go-to-market plan look like?	We will start with small dataset. We will try to improve our model by collecting more data. At first, we will deploy model on AWS or Google Cloud. We have a plan to deploy in local in the future. As we know it is a government service to verify vehicle registration, this product will only use the government selected officers who are in charge for verification. We will take feedback from them and try improve the best way.

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?	We will continuously add new images to improve performance by training. We will analysis the customer review and the success metrics to give them with a better performance and experience. We will be adding balanced data which cover every Bengali digits. We will also use quality images. Active learning will be the main part of this product as we will be getting new data from users.
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	To mitigate bias, we will be acquiring data from various arc and combination of numbers to keep our database balanced. Active learning will be one of the great factors to mitigate bias because it will continuously be added new images with labeling. We always need a human to check the performance and monitor everything.