

# Capstone Project Proposal



*ATTARY achraf*

---

## Business Goals

<p><b>Project Overview and Goal</b></p> <p>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?</p>	<p>In a business we receive a great number of letters and reading all those letters is Tedious and take a lot of time and resources. This is developing a program that will classes the letters after being scanned to tell us which letters are more important and should be read/prioritized sound like a wise idea. Machine learning/artificial intelligence can help us realize this objective and able us to automize this prosses so that we can focus the energy and time once used in manually reading and seeing whether a letter is important/urgent or just another advertisement or non-urgent administrative letters in something else</p>
<p><b>Business Case</b></p> <p>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</p>	<p>This is a very important problem to solve because it is going to help us relocate the resources (personnel) that used to do this task manually, this will help us save budget in a way that the employee will now do more valuable work like helping a team realize a project. This will also allow us to</p>

and/or other drivers of business success.	find the good letters which often include good business opportunities faster and thus maybe make deals faster than our competitors (in case of a new supplier that send multiple letters to different companies to find new prospect)
<p>Application of ML/AI</p> <p>What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?</p>	<p>The machine learning will have to</p> <p>1_detect the text in a scanned paper</p> <p>2_give a degree of importance to each letter (from 1 to 5,</p> <p>1 being irrelevant advertisement, 5 being urgent letters like urgent legal letter or new business opportunities.</p>

## Success Metrics

<p>Success Metrics</p> <p>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.</p>	<p>The business metric that we will use is the time the affected employee will have reading the letters. If the employee spends less time reading the letter as an effect of eliminating irrelevant letters and classifying the letter in a matter of importance we will consider that this initiative was a success</p>
--	--

## Data

<p><b>Data Acquisition</b></p> <p>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?</p>	<p>Data is freely available since all businesses have a huge amount of letter in their archive, those letters are already somewhat labeled witch will make the labeling work easy. Data is already available and will continue to be available as more letters come in every day. We will use about 500 letters,100 of each category.</p>
<p><b>Data Source</b></p> <p>Consider the size and source of your data; what biases are built into the data and how might the data be improved?</p>	<p>Since the useless letters like advertisement and irrelevant letters tend to be thrown away instead of archived the data will be biased toward the most important letters since they are the ones that are most likely to get archived.to improve the data we should keep the non-important letters until we have a clean and balanced data before we get rid of them. Since the data used to train the model will be of our own company there will be no concern of privacy/personally identifying information.</p>
<p><b>Choice of Data Labels</b></p> <p>What labels did you decide to add to your data? And why did you decide on these labels</p>	<p>We will use a 0-5 scale label.</p> <ul style="list-style-type: none"><li>1: irrelevant letters</li><li>2: non important non urgent</li><li>3: non important urgent</li></ul>

versus any other option?	<p>4: important non urgent</p> <p>5: important and urgent</p> <p>We choose those labels because they directly influence the choice made of what to do with the letters</p>
--------------------------	--

## Model

<p><b>Model Building</b></p> <p>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?</p>	<p>We will use in-house team to build the model since the letter may contain sensitive or secrete data that must not be revealed to third parties</p>
<p><b>Evaluating Results</b></p> <p>Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?</p>	<p>The model should, in case of doubt always put a letter above the level that he this is appropriate, this is to avoid having important letters labeled as non-important and make us loose opportunities this is why precision should be as high as possible (precision will be prioritized on recall). So we will consider that our model in functional when we get to an f1-score of 75% and precision &gt; 80 %</p>

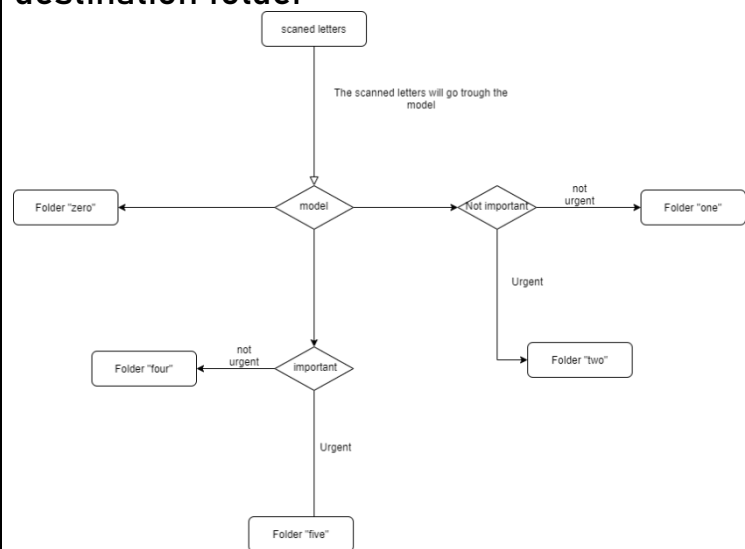
# Minimum Viable Product (MVP)

## Design

What does your minimum viable product look like? Include sketches of your product.

The minimum viable Product will only took the scanned letters and put theme in different folders (named one, two , three , four ,five ) according to what label will the model give theme

The minimum viable Product will not have any interface only a script and the destination folder



## 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 are designing this for our inhouse use at first. And when the model will become good enough for commercial use, we will use it put it in the market and it will be used by other companies as well that have the same problem and would like to optimize their human resources

	Who they are	What is their main goal	What is their main barrier to achieve this goal
Persona	General business and small companies that receive mails to	They want to allocate their human resources better and have faster	They have to assign an employee to manually go through all the

		their local	access to their important letters without having to go manually through all the letters	letters that they receive every day and they don't have to a way to automate this
Roll-out	<p>How will this be adopted? What does the go-to-market plan look like?</p>			
	<p>Target Market: the target market are businesses in general.</p> <p>Value Proposition and Product Messaging: We will help businesses automate the categorization process of the letters and permit theme to have quickly access to important letters.</p> <p>Pricing Strategy: the business model of our product will be as monthly subscription with reduced price to companies that will allow us to use their data to improve our model</p> <p>Distribution Plan : the model will be available via an API and all potential users will have a token that will permit theme to use a script as long as they are subscribed</p>			

## Post-MVP-Deployment

<p><b>Designing for Longevity</b></p> <p>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?</p>	<p>We will improve the model by continuously improving it using the data that we receive every day. And we will give great concern to keep the data balanced and clean so that we can have the best results. We will also perform manual testing some times by taking letters that we will label ourselves and see what the model will decide and with correcting the model when he mislabel a letter. We can employ A/B testing by the help of our close partners during the beta-phase by giving half the participants the model and the other not and see what improvement the group with the model will perceive</p>
<p><b>Monitor Bias</b></p> <p>How do you plan to monitor or mitigate unwanted bias in your model?</p>	<p>Since all letter will be eventually read except the one labeled 0.we will naturally correct the model as we go on. We should also perform some routines checks on the letters labeled 0 to see if everything is working as it should be, to do that, every day some random letter that have been rated as "0" will be manually verified ,by a designated employee to see if the model have labeled it correctly or not.</p>