

# Capstone Project Proposal



Svajune Klimasauskaite

## Business Goal

### 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.

### Project Goal

To help pupils or senior professionals to find a career path that best fit to them, i.e. based on the age, personality, experience and job market needs.

When the career path is chosen, then advice them with the online classes which helps to fill the missing knowledge gap.

### Problem

A career path is a difficult choice, that is not easy neither for the young people nor for the seniors. The current life pace shows that the future generation will be exposed to the career path change much more often than in the past.

Some of the youngsters, just after graduating the school, think they know exactly what profession they want to choose. Independent on the choice, some of them are getting lucky and become very good professionals simply by following what they love to do. While the others are left to struggle due to many reasons, such as, high competition, no market need, lack of soft skills, etc.

Due to digitalisation, senior professionals are facing the need to rethink their career path because of the change in the job market.

Nowadays there are tools and techniques that provide a self-paced education for anyone around the world. However, some of the education choices might bring more confusion than clarity and help.

This big variety of possibilities brings the question - "What better fits to me within my age, my personality, my skills and my country/city range".

### Why ML / AI

- ✱ ML / AI can help us to personalise the guidance for each individual.
- ✱ AI will help us to be in line with the job market needs, because of the 2 years prediction upfront.
- ✱ AI incorporates the personal assessment results, so that the person shall feel comfortable and curious within this area.

## 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.

## WHY

Even though the problem is the same relevant for the youngsters as for the seniors, the initial goal is to re-educate the senior professionals. And the reason for this is simply because we need to invite seniors to AI product development discussion table. Why?! - Because we need diversity, diversity, diversity !!!

Everyone can learn the ML/AI concepts, it is just the matter of time, money, will or interest. However it's not the case with the life practical experience, meaning you can never learn/understand until you experience on your own the childbirth, take care of your elder parents, lost of your beloved ones, the aging feel and etc...

## Impact

Nowadays, there is a profession called "Career coach" that gives similar services. However they are providing manual recommendations which still requires a lot of time of reading, finding, processing and most important understanding the information.

Here is a list of target groups and revenue strategy for every group:

- ✳ individuals. On the individual level we want to provide the career path recommendation with the best time and money ratio. Similar courses might cost 2 or 3 times more depending on the education school. By providing our payed services we will save a lot of clients money
- ✳ teams. When you are choosing a new team member for a new role you might not be aware about the quality standards for that position.
- ✳ organisations. This could be an alternatives for SAP Success Factor or Workday products (which are very expensive and not affordable for smaller companies).
- ✳ country. Regional Unemployment Office could use this product to bring people back to the job market in the most efficient way.

My focus during the product development will be on the individual level. If you change enough quantity of individuals, then they will change the team, organisation and eventually the country.

For product scaling we will come up with additional solutions for:

- ✳ organisations to do staff qualification assessment,
  - ✳ regional unemployment office to bring people back to the job market,
- which must boost the product revenue.

<p><b>Application of ML/AI</b></p> <p>What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?</p>	<p><b>AI Tasks</b></p> <ul style="list-style-type: none"> <li>✳ Personalise the career guidance for every individually, based on the life experience, education, years of professional experience, industries, technologies, responsibilities, achievements, etc.</li> <li>✳ Monitor the local job market in order to come up with ~2 years predictions.</li> <li>✳ Analyse the personality based on the soft skills assessment.</li> <li>✳ Combine experience, personality and job market tendencies in order to offer the best career path fit.</li> </ul> <p><b>Business Outcome</b></p> <p>Personalised education program for the specified location which was prepared based on the client education, experience, personality, budget, time and job market tendencies.</p>
--	---

## Success Metrics

<p><b>Success Metrics</b></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><b>Success Metrics</b></p> <p>For the initial phase we should measure <i>if the AI Product solves the problem</i>. We can do this by measuring:</p> <ul style="list-style-type: none"> <li>✳ User enrolment ratio.</li> <li>✳ Tracking education progress.</li> <li>✳ User input through the questioners within various interaction with the product.</li> <li>✳ ~6 month after finishing re-education program ask the user for the status update.</li> <li>✳ Verify If we correctly understood the user expectation with the word “satisfaction”: <ul style="list-style-type: none"> <li>✳ Financial success,</li> <li>✳ Occupational success,</li> <li>✳ Flexibility,</li> <li>✳ Overall satisfaction, etc.</li> </ul> </li> </ul>
---	---

# Data

<b>Data Acquisition</b>  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?	<b>Public Data</b> <ul style="list-style-type: none"><li>* Job Ads</li><li>* Online Education Schools</li><li>* Universities that provide on-line education programs</li></ul> <b>Private Identifying Information (PII)</b> <ul style="list-style-type: none"><li>* Extensive CV</li><li>* Personality assessment results</li></ul> <b>Data Sourcing Strategy</b> <ul style="list-style-type: none"><li>* Public, dynamic information that has the daily change frequency shall be loaded through the APIs on a daily basis.</li><li>* Public dynamic information that doesn't have daily change frequency shall be loaded through the APIs on a monthly basis.</li><li>* In case the APIs are not available, then source through the flat file or web scraper.</li><li>* All data classified as PII is accepted as per the contract agreement and loaded through the GUI.</li></ul> <b>Data size</b> <ul style="list-style-type: none"><li>* The initial goal is to have 100 data set for every data class.</li><li>* However for some of the classes might be difficult to accomplish this initial goal.</li><li>* There is no general solution for the lack of data, therefore depending on the class we will have to come up with one or other data "hacking" ways.</li></ul>
<b>Data Source</b>  Consider the size and source of your data; what biases are built into the data and how might the data be improved?	<b>Bias Control</b>  Bias control is the key task since all the data sources have its own biases, e.g.: <ul style="list-style-type: none"><li>* Job Ads lean more to IT jobs</li><li>* Education options lean more to IT education.</li><li>* Bias from the CVs might be on the age group or gender.</li></ul> The important thing is to be aware of the bias and understand the impact on the model. Bias can be mitigated with the control of the data that is used for training the model.

<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 versus any other option?</p>	<p><b>Labels</b></p> <p>There are two types of labels:</p> <ul style="list-style-type: none"> <li>✱ Personality label</li> <li>✱ Profession label</li> </ul> <p>Personality label will have all the categories that are defined in the personal assessment test.</p> <p>Profession label will have all the categories that are defined within the education material.</p> <p><b>Why These Labels</b></p> <p>These 2 labels are the key product drivers, because <u>we want to identify the personality and find the best profession fit as it's described by the "book" and not by the job ad.</u></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><b>Model Building Strategy</b></p> <p>For the success of the product development, we need to be ready for the agility and flexibility, therefore we shall use an in-house team.</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>	<ul style="list-style-type: none"> <li>✱ F1 score will be used to measure the overall model performance.</li> <li>✱ In this business case, Precision is more important than Recall.</li> <li>✱ The level of performance will be defined in the Product Designing phase, when more detailed specification is available.</li> <li>✱ Additionally, we will create a quantitative measure called <u>Distance</u>, that shall assess how close our model mimics a Subject Matter Expert recommendation.</li> </ul>

# Minimum Viable Product (MVP)

<b>Design</b>  What does your minimum viable product look like? Include sketches of your product.	<p>GUI is very simple since all the complexity is in the model and admin console.</p> <p><b>User Input</b></p> <p>We'll ask for the 3 inputs from the user:</p> <ul style="list-style-type: none"><li>✱ Add your CV,</li><li>✱ Start personality assessment,</li><li>✱ Fill out the questionnaire.</li></ul> <p><b>Output</b></p> <p>Based on the input we'll generate several career paths with assigned job description that is applicable for this path. The user can go deep and check more features like required time for education, price, selected courses and other information.</p> <p><b>Mockups</b></p> <p>The following mockups are included:</p> <ul style="list-style-type: none"><li>✱ Website menu with available features</li><li>✱ Home page / Dashboard with analytics</li></ul>
<b>Use Cases</b>  What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?	<p><b>Epic</b></p> <p>The persona is male with 20 years of professional experience. Through out of his professional experience he changed employees 2 times. The industry he's working in is heavily regulated. The organisation was heavily hierarchical structure, so he used to work on the task level responsibility and be valued by the quality of task execution.</p> <p>Now his company is facing a strong competition on the global level as well as struggling with attracting young talents. Organisation goes for restructuring and adapts new approaches as Holacrasy, Agile, Flat structure, etc. in order to be more efficient.</p> <p>Restructuring includes internal trainings on the mass scale, with the expectation to have a completely retrain workers within ~2 months. When management expectation fails, then the new highly expensive manager is hired to fix it, who most likely chooses a quick and dirty solution. And in this way the employers that knew the organisation/business the best are lost.</p> <p>The "quick and dirty solution" could be mitigated if the individuals would have a tool to monitor their qualification according to the market needs.</p>

<p><b>Roll-out</b></p> <p>How will this be adopted? What does the go-to-market plan look like?</p>	<p><b>Go-to-market Plan</b></p> <p>The product is very complicated and it is clear that it's impossible to go to the market with the full and complete product, therefore we'll break it by profession and location.</p> <p>We will create a small product for several professions and one location. Incrementally we will source and train the model with more data from more locations and professions.</p> <p>The professions and location for the roll-out will be chosen based on the market and revenue opportunities.</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>This product can only be improved with</p> <ul style="list-style-type: none"> <li>✱ More time,</li> <li>✱ More client feedback,</li> <li>✱ More training,</li> <li>✱ More data.</li> </ul> <p>The importance is to monitor the product usability and identify areas that:</p> <ul style="list-style-type: none"> <li>✱ require fixes,</li> <li>✱ require usability improvements,</li> <li>✱ require UI explanations,</li> <li>✱ require A/B testing for understanding the user preferences,</li> <li>✱ show new business opportunities, etc.</li> </ul>
<p><b>Monitor Bias</b></p> <p>How do you plan to monitor or mitigate unwanted bias in your model?</p>	<p>Some of the classes will have more data, than the other. For the bias mitigation we can apply the data selection logic for the classes where we have lots of data.</p> <p>As well, we need to understand the bias source:</p> <ul style="list-style-type: none"> <li>✱ Annotation Bias</li> <li>✱ Data Bias</li> <li>✱ Model Bias</li> </ul>