**Machine Learning Proof of Concept (POC):**

https://www.linkedin.com/pulse/guide-launching-your-first-machine-learning-poc-without-shadi-naguib/

* Step 1 - Start with a Problem:
  + First, likely to start with big problem.
    - Example: Shortage of radiologist leads to delayed and inaccurate diagnosis.
  + Narrow down the problem
    - Example: focus on pneumonia in children instead of all diseases.
  + Build business case after defining problem and confirming with team it is worth solving.
    - Is it an impactful problem?
    - Can you quantify the business outcome?
    - Does the problem have substantial volume of associated data?
  + Validate feasibility.
    - How much data is needed?
    - Does the dataset effectively address problem?
    - Is the dataset complete and properly annotated?

**Example case Study:** Let's say we agree that **a tool for doctors capable of rapidly and accurately identifying pneumonia cases** will enable early intervention and appropriate treatment. Even though we agree that this is a tool to support radiologists not replace them, it's still clear that it will not only improve patient outcomes and save lives but also streamline the diagnostic process and optimize resource allocation, creating significant business value.

* Step 2- Prototype and test with real data
  + Dataset preparation
    - To train your model, you need a dataset consisting of pairs of input and actual labels. Data must be capable of answering the questions you want your model to address.
      * Does the dataset align with a problem?
      * Is the dataset complete?
      * How can we annotate and ensure quality dataset overtime?
    - *Data availability* – Ensure there is an available dataset for your potential solution.
    - *Data Size* – Employing deep learning algorithms, data size becomes even more crucial.
    - *Data Distribution and Patterns* – To avoid bias, you need to take steps to address the imbalance in the datasets.
    - *Data annotation* – Convert unlabeled data into labeled data to prepare out datasets.
  + Model training
    - Here is where we input the data to train the Machine learning algorithm. Which Machine learning algorithm will be implemented?
    - Understand & diversify your data
      * + The quality of your training data determines the behavior and performance of your model. Poor data leads to a poor model.
        + For supervised learning models, the model can only learn what is present in the data. So, make sure all the necessary information is included.
        + Your data should cover all likely scenarios. In our case, we cannot expect the model to predict pneumonia cases without diaphragm shadow if we haven't provided images of such cases.
        + The provided data should represent real-world situations, not just artificially staged images. For example, if you're working with audio data, it should include background noise.
        + Ensure that your model receives an equal number of different types of data to avoid bias.
  + Model evaluation
    - Compare predicted labels against withheld actual labels.
      * **Accuracy = (True Positives + True Negatives) / (Total Samples)**
      * **Precision = True Positives / (True Positives + False Positives)**
      * **Recall = True Positives / (True Positives + False Negatives)**
      * [F1score](https://www.linkedin.com/feed/hashtag/f1score) is a combined metric that considers both precision and recall providing a balanced evaluation of a classification model's performance.  
          
        **F1 Score = 2 \* (Precision \* Recall) / (Precision + Recall)**
* Step 3 – Release. Learn, iterate.
  + AB Testing: Send 20% of customers to test new Machine learning and leave 80% with older version. Allowing you to see if the number support the efficiency of the new Machine learning.
  + Iteration: If new insight is gathered, create updated versions with updated annotations and labels.