**Framing a machine learning problem correctly**

Framing a machine learning problem correctly is crucial for achieving meaningful and effective results. Here’s a structured approach to framing a machine learning problem:

**1. Define the Problem Clearly**

* **What is the goal?**  
  Clearly articulate what you want to achieve. Example: "Predict the likelihood of a customer churning based on past interactions."
* **What is the business or real-world impact?**  
  Explain how solving this problem benefits the business, users, or research.

**2. Identify the Type of Machine Learning Problem**

* **Supervised Learning:** If labeled data is available.
  + **Regression:** Predict continuous values (e.g., house prices).
  + **Classification:** Predict discrete labels (e.g., spam or not spam).
* **Unsupervised Learning:** If labels are not available.
  + **Clustering:** Group similar items (e.g., customer segmentation).
  + **Dimensionality Reduction:** Reduce features while preserving information.
* **Reinforcement Learning:** If the model learns through trial and error.
  + Example: Training a robot to walk.

**3. Define Inputs (Features) and Outputs (Labels)**

* **What data is available?**
  + Identify the features (independent variables) that influence the target outcome.
* **What is the target variable?**
  + Clearly define what the model should predict (dependent variable).

**Example:**  
For predicting customer churn:

* **Features:** Purchase history, login frequency, customer support interactions.
* **Target:** Whether the customer churns (Yes/No).

**4. Determine the Evaluation Metric**

* **Classification Problems:** Accuracy, Precision, Recall, F1-score, AUC-ROC.
* **Regression Problems:** Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R² Score.
* **Clustering Problems:** Silhouette Score, Davies-Bouldin Index.

**Example:** If false positives are costly, prioritize Precision over Accuracy.

**5. Consider Constraints and Challenges**

* **Data Limitations:** Missing values, imbalance, bias.
* **Computational Constraints:** Real-time predictions, hardware limitations.
* **Ethical Considerations:** Fairness, privacy concerns.

**6. Establish a Baseline Model**

* Use a simple heuristic or traditional method (like linear regression) as a baseline.
* Compare advanced ML models against the baseline to measure improvements.

**7. Iterate and Improve**

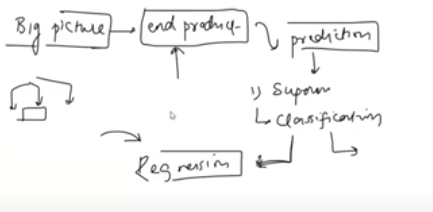
* Perform feature engineering, hyperparameter tuning, and model selection.
* Test different algorithms and evaluate their performance.
* Refine the problem statement if needed.
  1. **Business problem to ML Problem**

NETFLIX -> there is meeting how can we improve our revenue

* + 1. Using marketing get new customers
    2. Charge extra from existing customers
    3. Improve churn rate

Convert the business problem to mathematical model.

* 1. **Type of Problem.**

****

* 1. **Current Solutions**
  2. **Getting Data** 
     1. Watch Time
     2. Search but did not find
     3. Content left in the middle
     4. Clicked on recommendation ( order of recommendation )
  3. **Metrics to Measure**