Introduction:

With the success of ML in recent years, it has started attracting a lot of attention from the research and business communities. Machine Learning involves the design and development of pipelines for applications and ML systems.

Building such a pipeline requires a team of human experts: data scientists having statistical and ML knowledge; domain experts with years of experience within a specific domain. Together, these human experts can build a sensible ML pipeline containing data preprocessing, meaningful feature engineering, and fine-tuned models leading to great results.

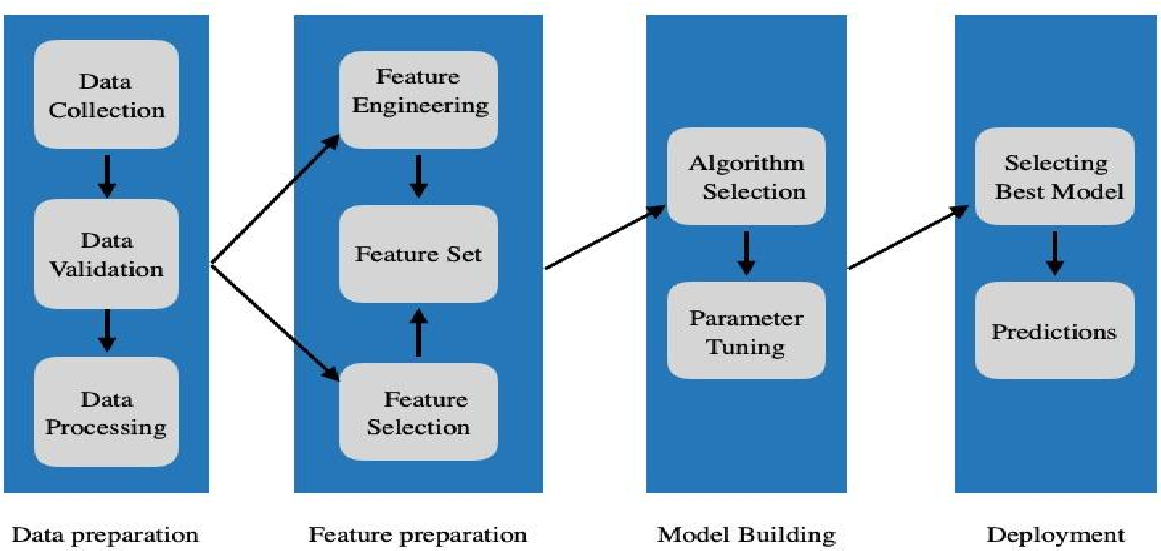
Every machine learning service, at its core, needs to solve the same fundamental problems: of deciding which machine learning algorithm to use on a given dataset, whether and how to preprocess its features, how to generate meaningful features, and how to tune all hyperparameters.

This process is a complex task, performed in an iterative manner with trial and error. Building a good ML pipeline is a laborious process and practitioners often use a suboptimal default ML pipeline. To solve these issues, a novel idea of automating the entire pipeline of machine learning (ML) has emerged, i.e., automated machine learning (AutoML). The goal of AutoML is to allow practitioners to build ML applications without much requirement for statistical and ML knowledge.

With the advancements in Cloud and BigData technologies, AutoML has started gaining a lot of attention. A complete AutoML system can dynamically combine various techniques to form an easy-to-use end-to-end ML pipeline system.

In the context of supervised learning, AutoML aims at removing user interaction altogether from all stages of the design and development of supervised learning systems. As data is being generated at an incredible rate in practically any context and scenario, the number of practitioners available to analyze such data are limited.

AutoML pipeline consists of several processes: data preparation, feature engineering, model generation, and model evaluation. Given below is an illustration of a typical ML pipeline.



References:

Automated Machine Learning - a brief review at the end of the early years: Hugo Jair Escalante

Efficient and Robust Automated Machine Learning: Matthias Feurer, Aaron Klein Katharina Eggensperger, Jost Tobias Springenberg, Manuel Blum, Frank Hutter

Automated Machine Learning: State-of-The-Art and Open Challenges: Radwa Elshawi, Mohamed Maher, Sherif Sakr

AutoML: A Survey of the State-of-the-Art: Xin He, Kaiyong Zhao, Xiaowen Chu

Benchmark and Survey of Automated Machine Learning Frameworks: Marc-Andr´e Z¨oller, Marco F. Huber