**Challenges in ML**

1. **Data Collection**. In life problem you have to fetch data from API or Web Scrapping.
2. **Insufficient Data / Labeled Data.**
3. **Non-Representative Data.** (Sampling Noise, Sampling Bias)
4. **Poor Quality Data.**
5. **Irrelevant Features.** (Garbage in, Garbage out)
6. **Overfitting.**
7. **Underfitting.**
8. **Software Integration.**
9. **Offline Learning / Deployment.**
10. **Cost Involved.**  (https://research.google/pubs/pub43146/)

**Real Life Applications of ML**

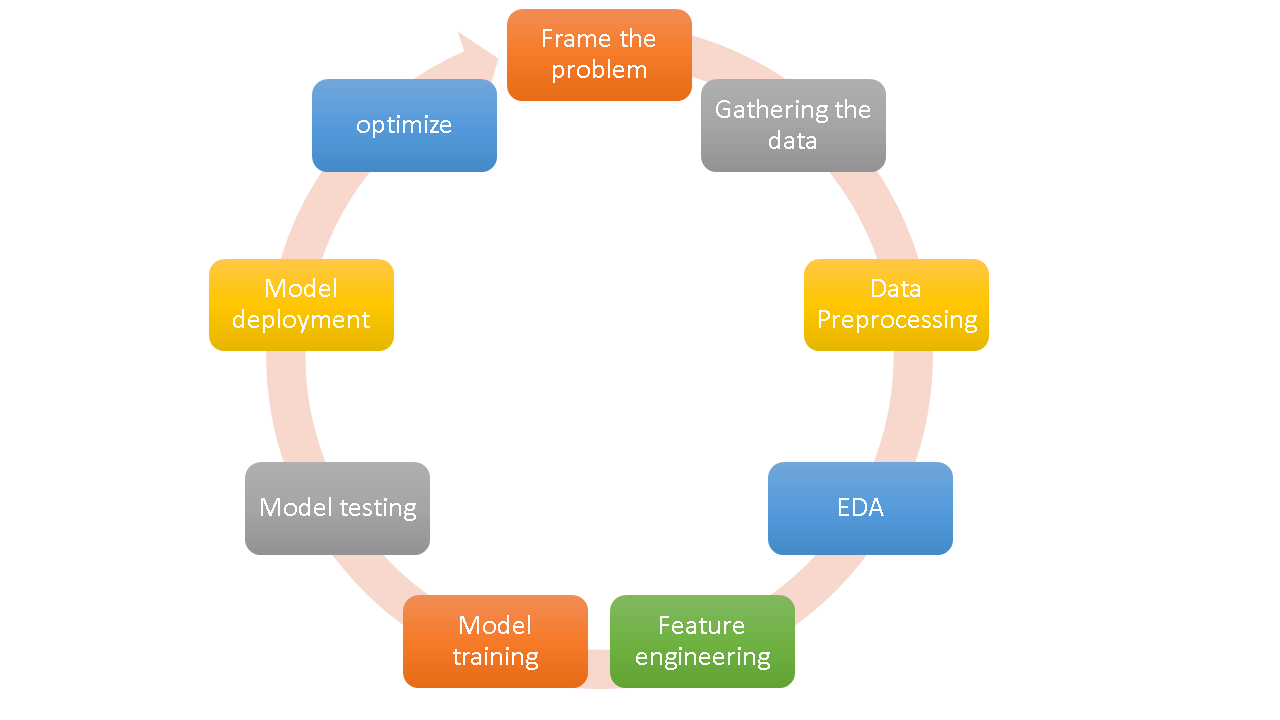
AI is already in our life this is not a new thing. i.e Facebook, YouTube, amazon and 100’s of examples.

I will give B2B examples not B2C because its so obvious.

1. Retail – Amazon / Big Bazar. If you are not paying for a product then you are a product. (FB, WhatsApp, Insta….)
2. Banking and Finance.
3. Transportation – Uber.
4. Manufacturing – Tesla. Predictive maintenance.
5. Consumer internet – Twitter. (Selling tweets for sentiment analysis)

**ML Development Life Cycle – MLDLC**

DLC are set of guide lines which you have to follow for developing any product. For Machine Learning product we have to follow the MLDLC. This is crucial for developing end to end products. Currently all the steps I am discussing are not defined in any book. But the core idea is same.



1. **Frame the problem**: In this step you gather the idea how to proceed further i.e selection of data, selection of model, cost of the product etc.
2. **Gathering the data:** Collecting and preparing data for use in training and testing a machine learning model. Data gathering from CSV, API, Web scrapping and storing it in correct form.
3. **Data preprocessing:** Cleaning, preprocessing, and transforming the data into a format suitable for training a model. Removing outliers, removing duplicate adding missing values are the essentials of data preprocessing.
4. **Exploratory Data Analysis:** EDA is a step that involves analyzing and visualizing the data to gain insights into its characteristics and patterns. This step helps to identify the relationships between the variables and understand the distribution of the data. EDA helps to identify the most relevant features for the model and helps in feature engineering. Do univariant and bivariant , outlier detection and imbalance to balance data.
5. **Feature engineering:** Feature engineering is a step that involves creating new features from the existing features to improve the performance of the model. Feature selection is a step that involves selecting the most relevant features for the model. Both feature engineering and selection help to improve the accuracy and performance of the model.
6. **Model Training, Evaluation, and Selection:** Model training is a step that involves building and training the machine learning model using the preprocessed data. In this step, you need to choose the appropriate algorithm and train the model on the selected features. Once the model is trained, you need to evaluate its performance using various metrics and select the best model.
7. **Model Deployment:** Model deployment is a step that involves deploying the machine learning model into production. In this step, you need to integrate the model with the existing system, create an API for the model, and deploy it on a cloud-based platform.
8. **Testing:** Testing is a step that involves testing the deployed model to ensure its performance in real-world scenarios. In this step, you need to test the model for various edge cases and ensure its robustness. If any issue arises here and the model doesn't seem fit for the requirements, you might need to iterate the steps again from start to accommodate the requirements.
9. **Optimizing and Improving the Model:** Optimizing and improving the model is an ongoing process that involves monitoring the model's performance and making improvements to enhance its accuracy and performance. This step involves retraining the model with new data, tweaking the hyperparameters, and fine-tuning the model to improve its performance.

**Data Engineer Vs Data Analyst Vs Data Scientist**

**Data Engineers** are the intermediary between data analysts and data scientists. As a data engineer, you will be responsible for the pairing and preparation of data for operational or analytical purposes. A lot of experience in the construction, development, and maintenance of the data architecture will be demanded from you for this role. Usually, in this role, you will get to work on Big Data, compile reports on it, and send it to data scientists for analysis.

**A Data Scientist** employs advanced data techniques such as clustering, neural networks, decision trees, and the like for deriving business insights. In this role, you will be the senior-most in a team and should have deep expertise in machine learning, statistics, and data handling. You will be responsible for developing actionable business insights after they get inputs from Data Analysts and Data Engineers. You should have the skill-set of both data analyst and data engineer. However, in the case of a data scientist, the skill sets need to be more in-depth and exhaustive.

**A Data Analyst** occupies an entry-level role in a data analytics team. In this role, you need to be adept at translating numeric data into a form that can be understood by everyone in an organization. Moreover, you need to have required proficiency in several areas, including programming languages such as python, tools such as excel, fundamentals of data handling, reporting, and modeling. With enough experience under your belt, you can gradually progress from a data analyst to assume the role of a data engineer and a data scientist.

|  |  |  |
| --- | --- | --- |
| **Data Analyst** | **Data Engineer** | **Data Scientist** |
| Data Analyst analyzes numeric data and uses it to help companies make better decisions. | Data Engineer involves in preparing data. They develop, constructs, tests & maintain complete architecture. | A data scientist analyzes and interpret complex data. They are data wranglers who organize (big) data. |

**Skill-Sets**

The below table illustrates the different skill sets required for Data Analyst, Data Engineer and Data Scientist:

|  |  |  |
| --- | --- | --- |
| **Data Analyst** | **Data Engineer** | **Data Scientist** |
| Data Warehousing | Data Warehousing & ETL | Statistical & Analytical skills |
| Adobe & Google Analytics | Advanced programming knowledge | Data Mining |
| Programming knowledge | Hadoop-based Analytics | Machine Learning & Deep learning principles |
| Scripting & Statistical skills | In-depth knowledge of SQL/ database | In-depth programming knowledge (SAS/R/ Python coding) |
| Reporting & data visualization | Data architecture & pipelining | Hadoop-based analytics |
| SQL/ database knowledge | Machine learning concept knowledge | Data optimization |
| Spread-Sheet knowledge | Scripting, reporting & data visualization | Decision making and soft skills |

For more info visit = https://medium.com/campusx/exploring-the-mystery-behind-different-job-titles-for-machine-learning-and-data-science-9352849f283a