Data

Data Prerequisites

1. For ML, training data with input and output is required for supervised learning
2. All data encoded as numbers
3. Typically High dimensional

Data Curation

1. Open Source: Data available online (Google AI, data.gov.in, UCI Machine Learning Repository)
2. Crowdsourcing: If you are rich, outsource data-gathering online (amazon mechanical turk, dataturks, figure eight)
3. Self work(You are smart): You can create your own data for training.
4. Wikidata: Has a lot of hindi-english translations