**Introduction**

* Today we will be going through intro to ML.
* We will start from the basics and theory, and then move towards implementing the algorithms in python
* So, by the end of this session you would be able to code ML models and use them for various real-life applications
* So here is the overview

**AI, ML & Data Science Definitions**

* This slide is a myth buster
* People get confused by these 3 terms and use them interchangeably
* AI – idea of what intelligence in machine should be
* ML – one way of achieving AI
* Data Science – culmination of various fields

**Traditional programming vs ML**

* Let’s take example of predicting house prices in a particular area, given the area of the house and other params like crime rate, climate etc.
* In **TP,** you have to come up with your own formula with ties all the things together, for instance housing price = 1000$ x area sq ft or something like that, which is a difficult task. Other example may include classifying image into different categories.
* Challenges: hard coded hence coming up with the universal magical formula is impossible for a task.
* This is where TP fails and ML comes into picture, In ML you examine the data and use the right ML algo’s. Then use data to train the machine.
* Remember: ML is highly dependent on data

**Machine learning**

* Fun fact: ML is not new, in fact it’s old and was conceptualized in 1950. Like Bayesian methods.
* But what changed then? due to culmination of things like:
* cheap storage, resulting in rise in data
* Computation power – more complex models
* Are some of the main factors for the rise in use of ML

**Fields of application**

* Computer vision: visually challenged camera, mobile camera, Self-driving cars
* Speech recognition: assistants like – Siri, Alexa, google assistant
* Financial analysis: companies use FA to predict stock prices - zeroda
* Search engines: google- content suggestion
* Ads-targeting: fun fact fingerprinting- business model of YouTube and Facebook, suggest content for max user engagement, controversial topics etc.

**Machine learning Pipeline**

* Pipeline is a series of tasks/processes that take place in between the starting point to reach the completion
* So, ML pipeline is the sequence of tasks or methods used to go from raw data and using it to get predictions for the task at hand

**Ingestion**

* First step, which involves gathering of data

**Preparation**

* Ingested Data is not always clean, for instance in image processing – OCR
* There can be blurry text in noisy background, so we have to use techniques like dilution and erosion to get a clear and highlighted text in all the images

**Retraining**

* Most important step, for machine to give relevant and correct prediction in the future