Types of Machine Learning

Day 5: Online ML

⇒ River, Vowpal Wabbit for online-ml (Python libraries)

- Learning rate: Choosing how frequently to train

- Out of Core Learning: Use online learning on potato. It's like doing the offline training (initial) in batches (continue style)
  - ↳ Tricky to use
  - ↳ Risky
  - ↳ Hard to rollback
  - ↳ Evals and data integrity checks

- Day 6: Instance Based vs Model Based
  (memorize) (generalizing)

- Instance Based ⇒ K Nearest Neighbours → RBF, Kernel Network
  (no training or learning) But need data in inference

- Decision Function / Boundary.
  Don't need data in inference

Problems in ML

Day 7:

1. Data Collection
   - API or Webscraping

8. Insufficient / Labelled
- When you have large data, algorithm won't matter much.

Non Representative Data
- The Data doesn't represent whole situation or is biased.
eg. survey winning country of cricket only in India
- This is called 'sampling noise'
- Even though you collect all data there is still hidden bias ⇒ 'sampling bias'
(Nationality of other ppl in Country)

Poor Quality Data
Irrelevant Features
Overfitting
Software Integration

9. Real-life Applications (B2B)

Day-8: Retail (Amazon)
Association Rule Based Learning (with products bought together)
Banking
Correlation with defaulters
Transport
Surge Pricing
Manufacturing
Predictive Maintanence
X.com / Consumer Internet
Sentiment Analysis
↳ Hmm not so accurate (edge case)
Election Prediction
↳ Sell to stock ppl like JP Morgan

M T W T F S S

Page No.:

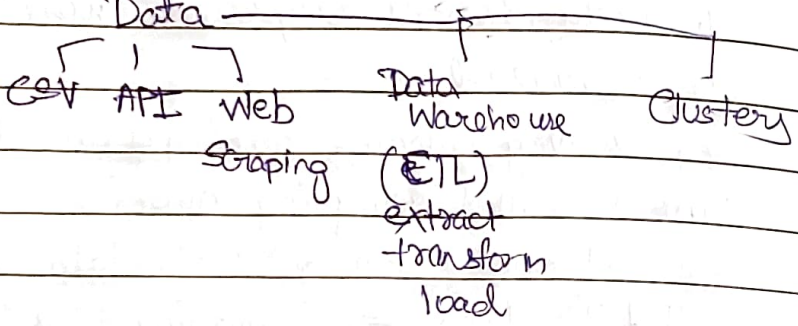
Date:

YOUVA

TMP in practice

Day-9 Machine learning Development life Cycle

1. Frame the Problem
2. Gather Data



3. Data Preprocessing
   - Remove Duplicates, missing values, outliers, scaling (standardization)
4. Exploratory Data Analysis:
   - Vizualization, Univariate Analysis, Bivariate, Multivariate,
   - Outlier Detection.
   - Handle Imbalance
5. Feature Eng and Selection
6. Model Training, Evaluation and Selection
   - Hyper-parameter Tuning
   - Joining different models (Ensemble Learning)
7. Testing
   - Take feedback.
   - A/B Testing
8. Optimize
   - Backup, Data, Load balancing, Re-train

Page No.:
Date: YOUVA

Day-10 AngelList ⇒ Jobs in Startup

Data Analyst	Data Engineer
Data Scientist	ML Engineer

- One person doesn't do all 4 steps
- Engineer - Scrape Data from given source
  OLTP
  OLAP
  Move / Store data in warehouse
  Pipelines / API to access data
  Maintaining it.
- Analyst -
  Cleaning and Organizing Raw Data
  Analyze data to derive insight
  Creating data visualization
  Producing / Maintain report
  Optimizing collection procedures
- Scientist - Full Stack Guy, thinks about future
  ↳ use case (Predictive modelling)
  ↳ Main Goal
- ML Engineer - Small role
  Deploying to production, Scaling and optimizing for prod
  Monitoring / Maintenance

Day11: Tensors

- 0D Tensor / Scalar

Day12: Setup Day13: End-to-End Toy Project

P.T.O

Day 14: 1) Business Problem to ML.

- Stop leaving customers ⇒ churn rate (how many consumers leave per unit time)

2) Type of Problem

Big picture → find out who's leaving

↓ discount ↓ what issue? ↓ probability of leaving ∝ discount

So, it's a regression

Day 15: Data gathering with CSV