

Roadmap to Become a Machine Learning Engineer (MLE)

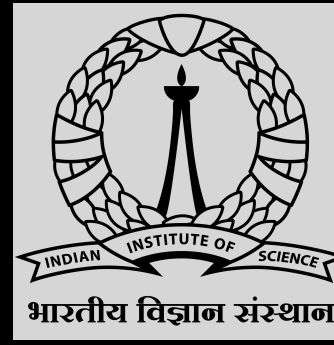
- ML Systems around us
 - ▶ Practical Examples
 - ▶ Need for ML solutions
- ML Pipeline
 - ▶ Tasks and Tools involved in the pipeline
 - ▶ Roles for each task
- Skills Needed to Become an MLE
 - ▶ Tasks involved in the pipeline
 - ▶ Tasks involved in the pipeline

About me - Ajay Shenoy

B. Tech - NITK, Surathkal



Ph.D - IISc, Bangalore



ML Coursework - Rigorous Math
Foundations

- Probability, Linear Algebra, Optimization

AI/ML - AlphaICs, Target, Harman



Optimising Deep Learning inference chip

ML algorithms in supply chain

Driver monitoring system

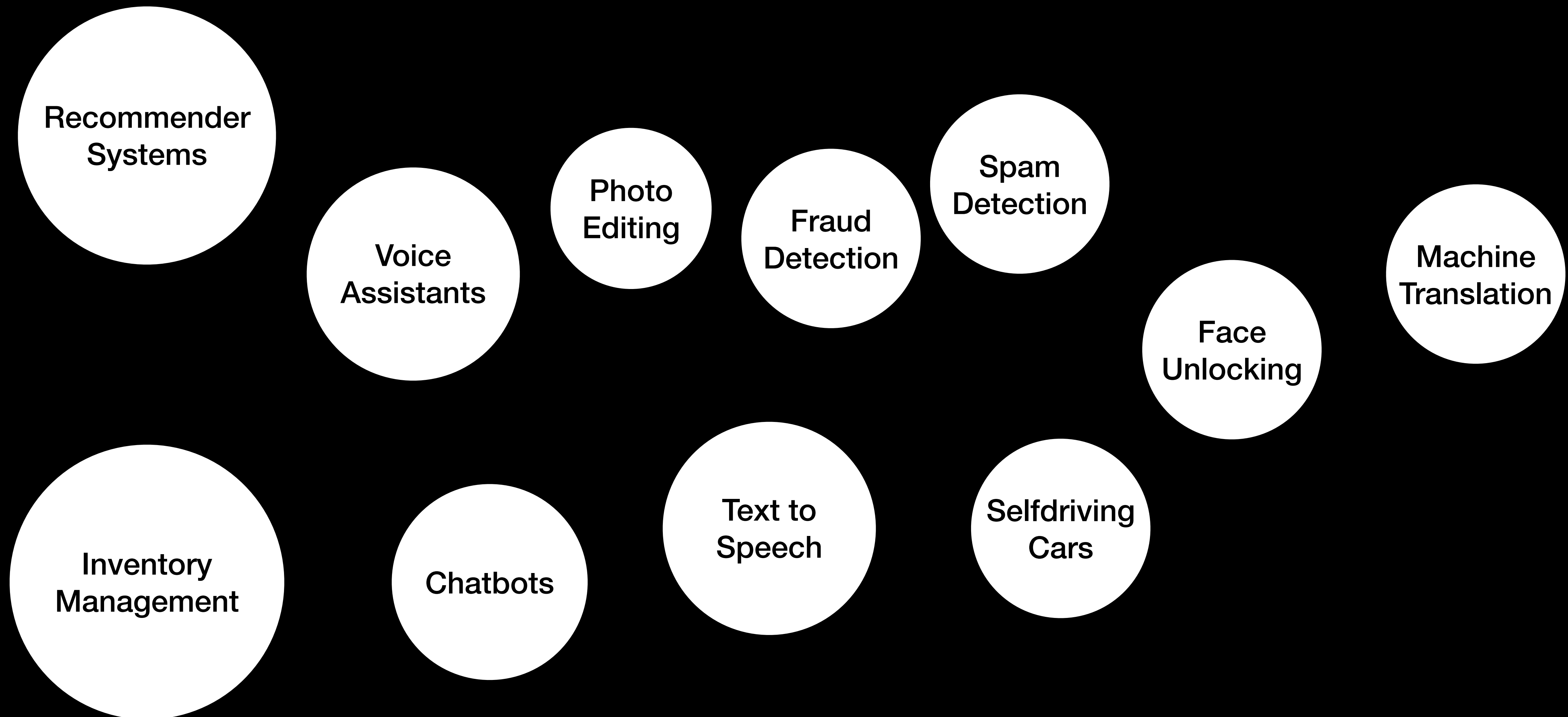
Instructor at Scaler



Taught about 2000 students at Scaler
We built structured DSML curriculum

LinkedIn <https://www.linkedin.com/in/bastyajayshenoy/>

ML Systems around us



ML Systems around us

- What do all these ML applications have in common?

Data ... and more data Tons of data!!!

This requires scaling our solution, and automating the process

- We need a solid **ML Pipeline**

Understanding this pipeline leads to a solid roadmap

Tasks, tools, skills

ML Pipeline

Data Ingestion

Typical Tasks:

- Collect data from various sources
- Ensure data is in a usable format
- Establish regular update

Tools: Pandas/SQL

Roles: Data Engineer

Data Validation

Typical Tasks:

- Check for missing or null values
- Identify and handle outliers
- Verify data types

Tools: Pandas/SQL

Roles: Data Engineer

Data Preprocessing

Typical Tasks:

- Handle missing data
- Normalize or standardize features
- Perform feature engineering

Tools: Pandas/Numpy/Scikit Learn

Roles: Data Scientist, Data Engineer

Model Training

Typical Tasks:

- Split data: training and validation sets
- Choose appropriate algorithms
- Set initial hyperparameters

Tools: Scikit Learn/Pytorch/Tensorflow

Roles: Data Scientist

Model Validation

Typical Tasks:

- Evaluate model performance
- Choose appropriate metrics
- Perform error analysis

Tools: Scikit Learn/SHAP/WandB

Roles: Data Scientist, ML Engineer

Model Deployment

Typical Tasks:

- Prepare model for production
- Set up API endpoints, CI/CD
- Implement logging and monitoring

Tools: Flask/Docker/Kubernetes/MLFlow

Roles: ML Engineer/DevOps Engineer

Skills Needed

Programming

- SQL
- Python
 - Numpy, Pandas, Matplotlib
 - Scikit Learn, Scipy, open cv
 - PyTorch, Tensorflow/Keras, Gym
 - Langchain, HuggingFace

ML Algorithms

- Classical ML (SKLearn)
 - Linear/Logistic Regression
 - Boosting (XGBoost)
 - Clustering
 - Dimensionality Reduction
- Deep Learning (Pytorch/Keras)
- Reinforcement Learning (Gym/RLib)

Data Analysis

- Data Cleaning
- Data Preprocessing
- Exploratory Data Analysis
- Feature Engineering

Software Engineering

- Software design principles and patterns
- Version control (Git/DVC)
- API development
- Containerisation (Docker)
- Deployment (AWS/GCP/Azure)

Mathematics

- Probability and Statistics
- Linear Algebra
- Calculus, Optimisation, Gradients

Resources

Scaler Topics: www.scaler.com/topics

Hacker Rank: hackerrank.com

Corey Schafer: <https://www.youtube.com/@coreyms>

Jake Vanderplas: <https://jakevdp.github.io/PythonDataScienceHandbook/>

ThinkStats: <https://greenteapress.com/wp/think-stats-2e/>

Sebastian Rashcka: <https://github.com/rasbt/machine-learning-book>

Software Engineering for Data Scientists

- <https://github.com/catherinenelson1/SEforDS>