Data Science in Action

Vaibhav Srivastav

Vaibhavs10@gmail.com | @vaibhavsriv10 | Github - Vaibhavs10

Image Attribution: Mike Braun, Chief Data Scientist, Zolando

Who am I?


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

The Audience (You!)

1. Know some Analytics/ Information Management already?
2. Know some Machine Learning already?
3. Both/ None of the above?

What is Data Science?

What is Data Science?

Data science is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining

Um, Example?

To find restaurants serving chinese food in gurgaon:

SELECT restaurant_name, restaurant_address FROM restaurants WHERE city = ‘gurgaon’ AND food_type = ‘chinese’

Natural way:

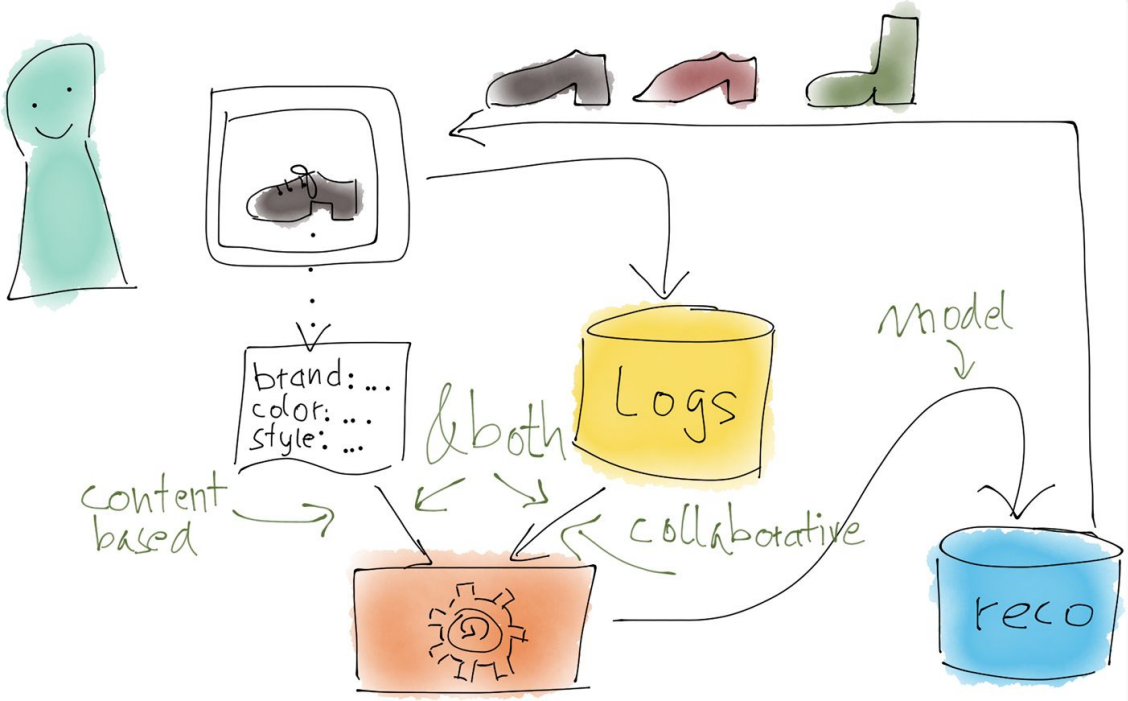
Where can I find some chinese food in gurgaon?

Case Study: Recommendation Systems

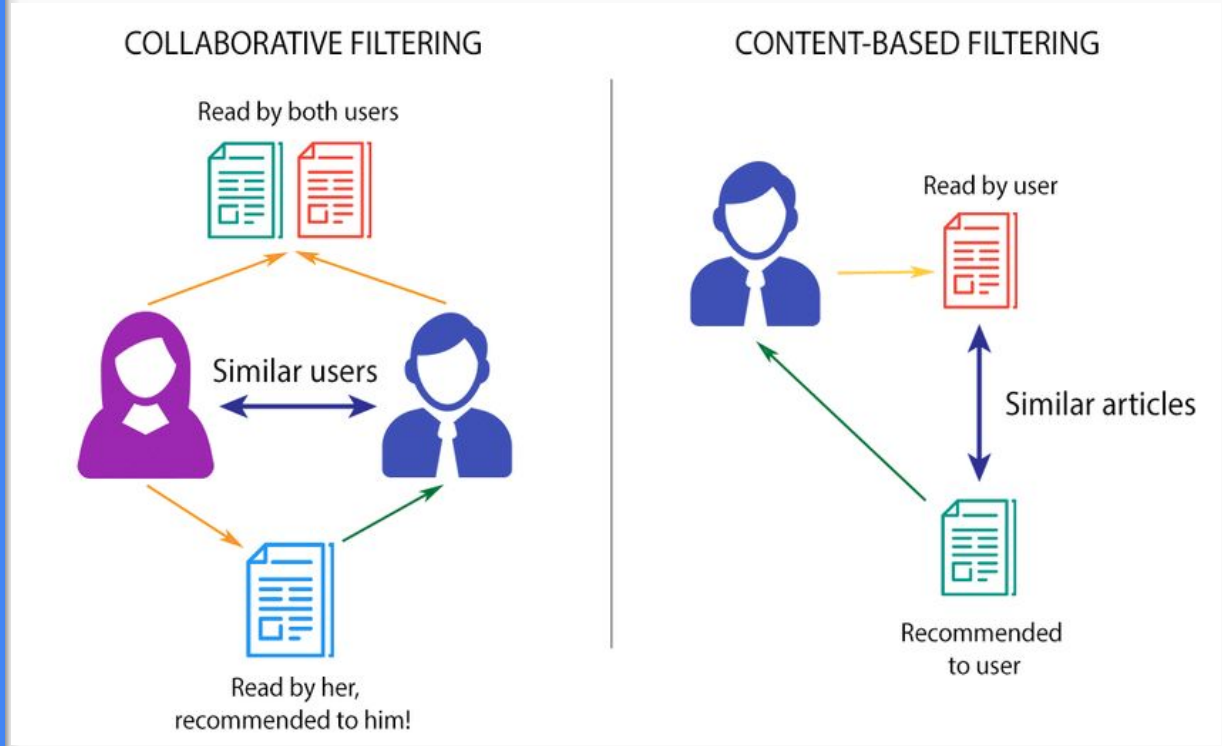
Expectations

1. Ask Questions!
2. Stop me if you don’t understand something
3. Relate with real world scenarios, not textbook

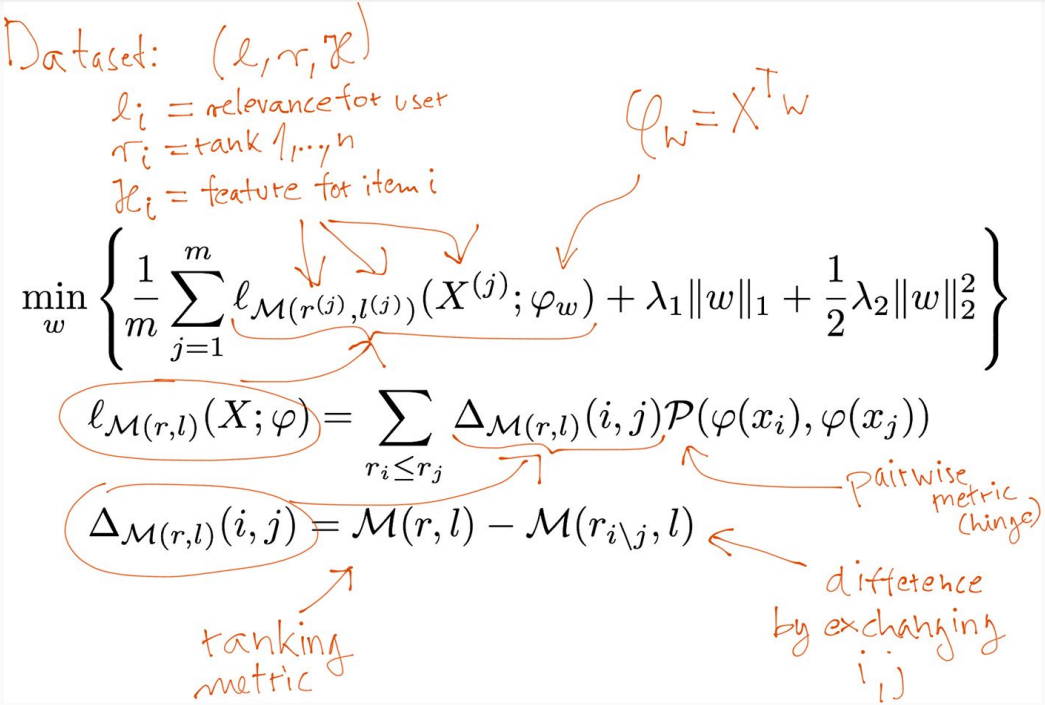
Problem Statement



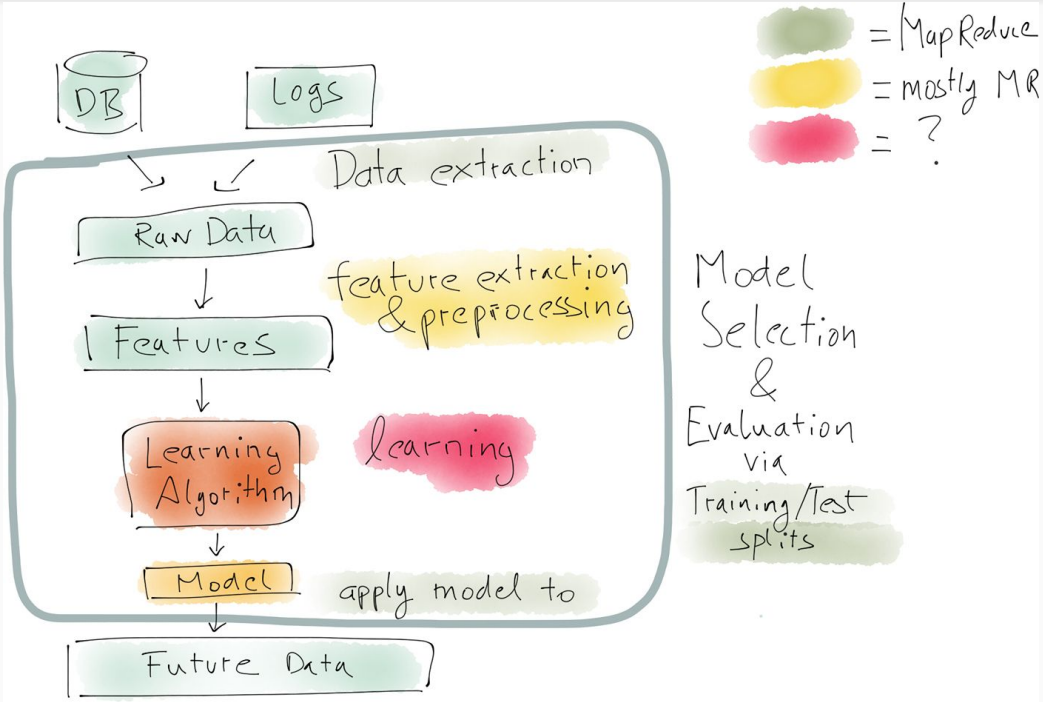
Content Vs Collaborative Filtering



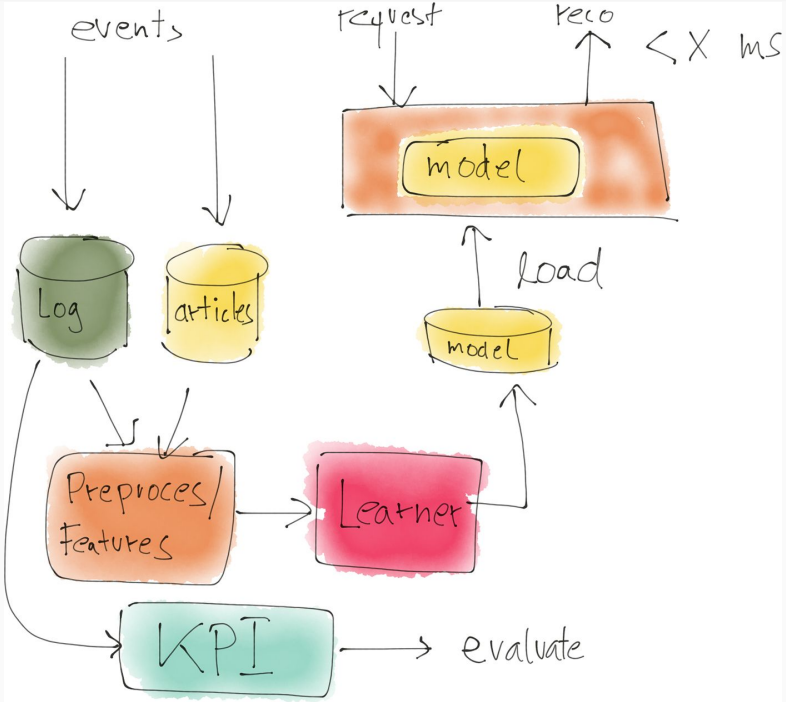
Behind the Scene: Math



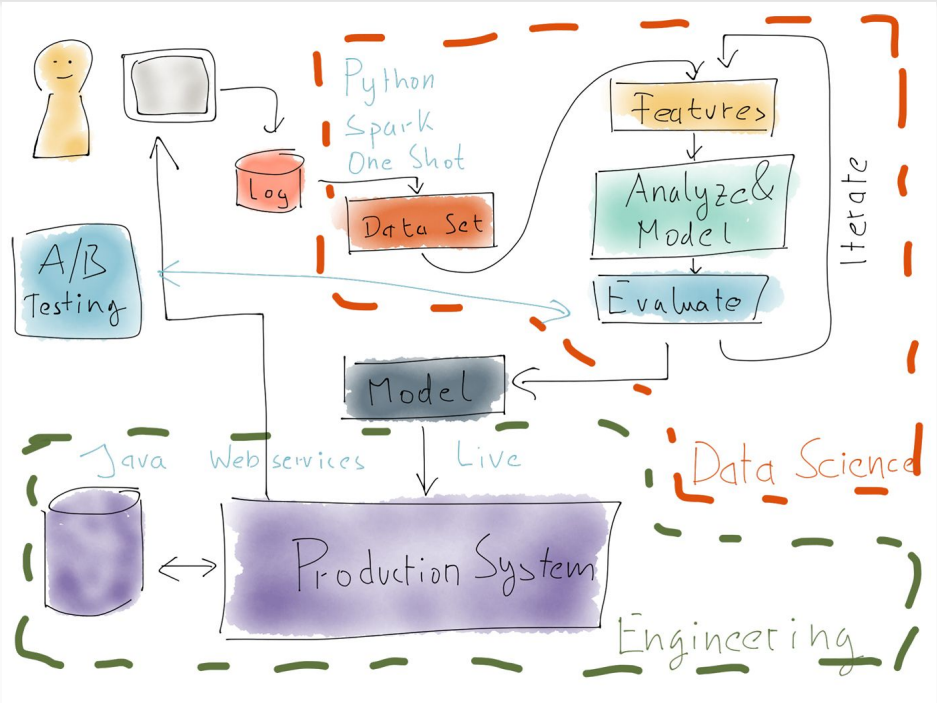
Analytics Setup



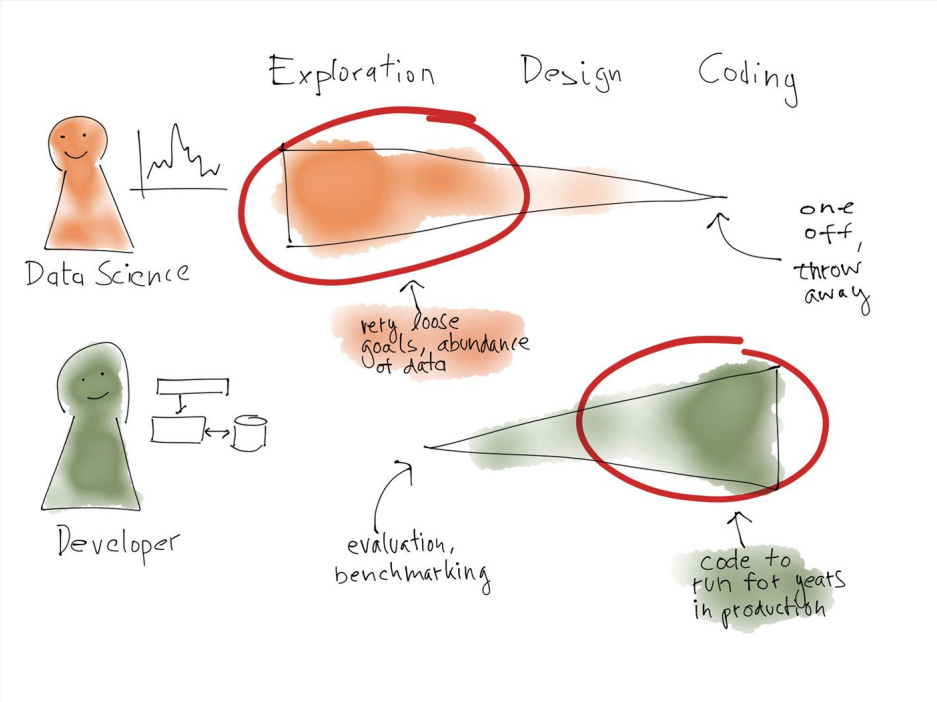
Model Serving Setup

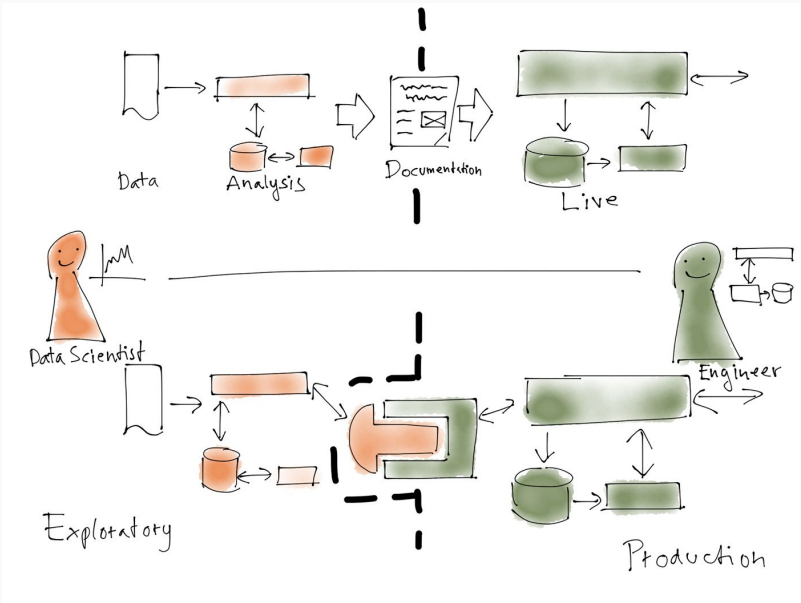


Big Picture


Python
Spark
One Shot
Log
Data Set
Features
Analyze &
Model
Evaluate
Iterate
A/B
Testing
Model
Java
Web services
Live
Data Science
Production System
Engineering

Data Guys


Exploration
Design
Coding
Data Science
one off, throw away
very loose goals, abundance of data
Developer
evaluation, benchmarking
code to run for years in production


Data
Analysis
Documentation
Live
Data Scientist
Engineer
Exploratory
Production

Key Takeaways

1. Know your data well
2. Draw out your output expectations (Regression, Classification, Recommendation)
3. Understand the problem statement
4. Start simple, always
5. Ask Questions as many as you can

Questions?