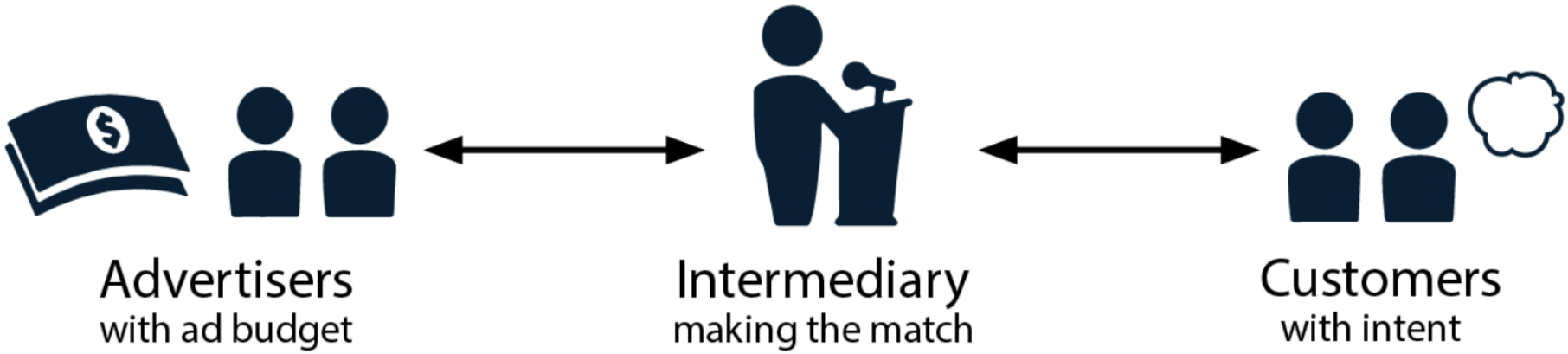
Large Scale Machine Learning in Digital Advertising

Seyed Abbas Hosseini
VP of AI, Tapsell Inc.
Assistant Professor, SUT
abbas@tapsell.ir

Outline

- Digital Advertising
  - Digital Advertising Ecosystem
  - Real Time Bidding
- CVR Estimation Problem
  - Problem Definition
  - Predictive Models
  - Production Challenges
- Tapsell Brain

Advertising



Conveying the message of the advertiser to the target audience

Traditional Advertising

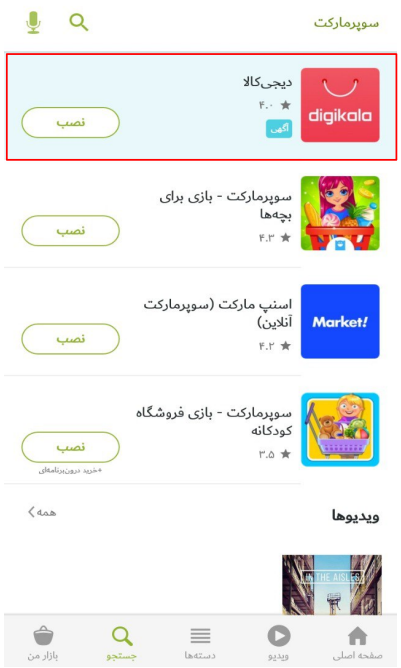
"Half the money I spend on advertising is wasted; the trouble is I don't know which half."

- *John Wanamaker*

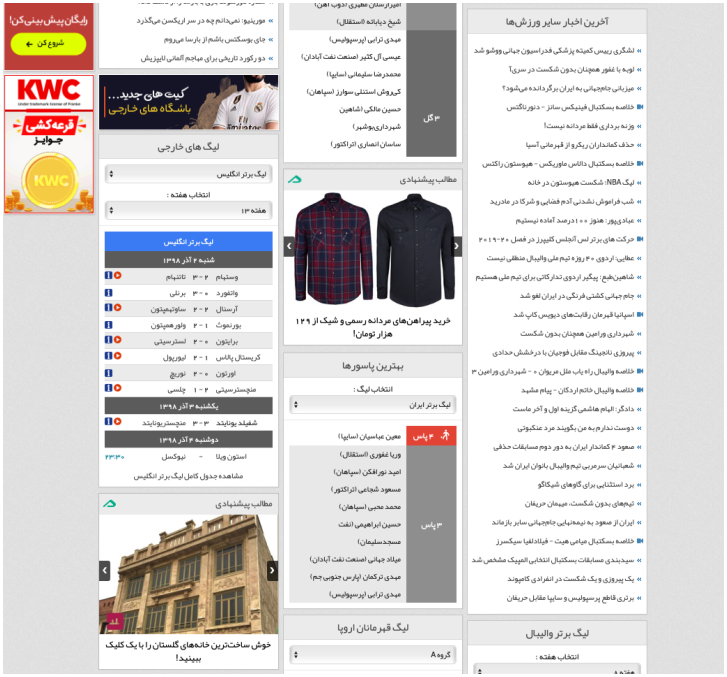
(1838-1922)

Father of modern advertising and a pioneer in marketing

Digital Advertising



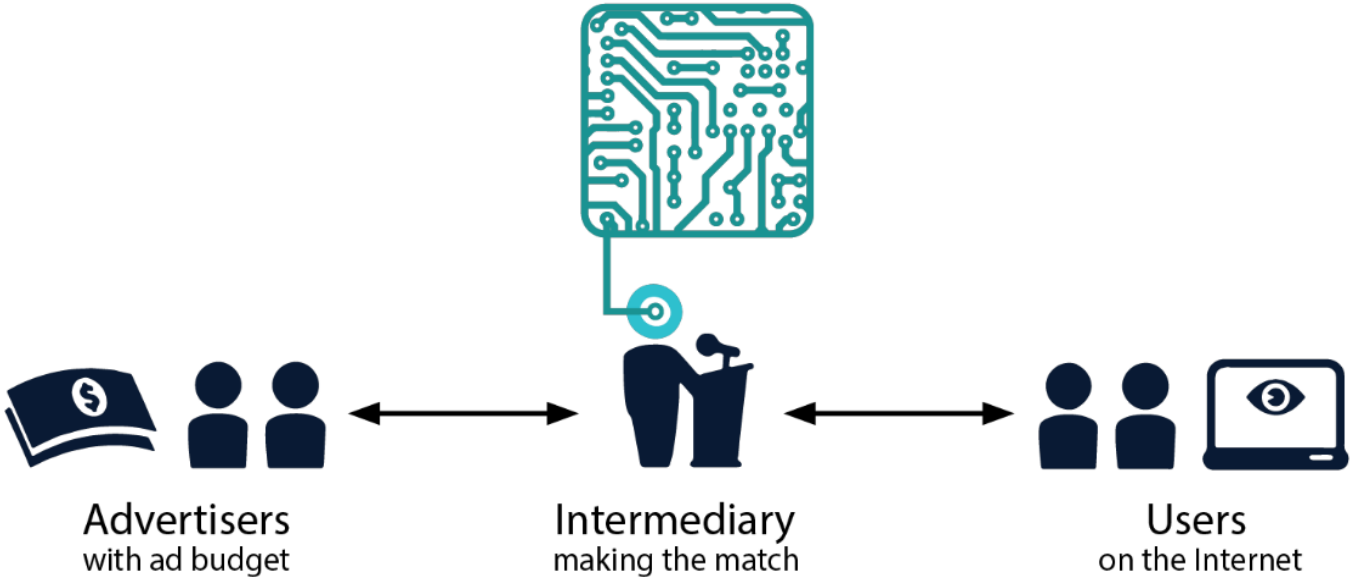
Sponsored Search



Display Advertising

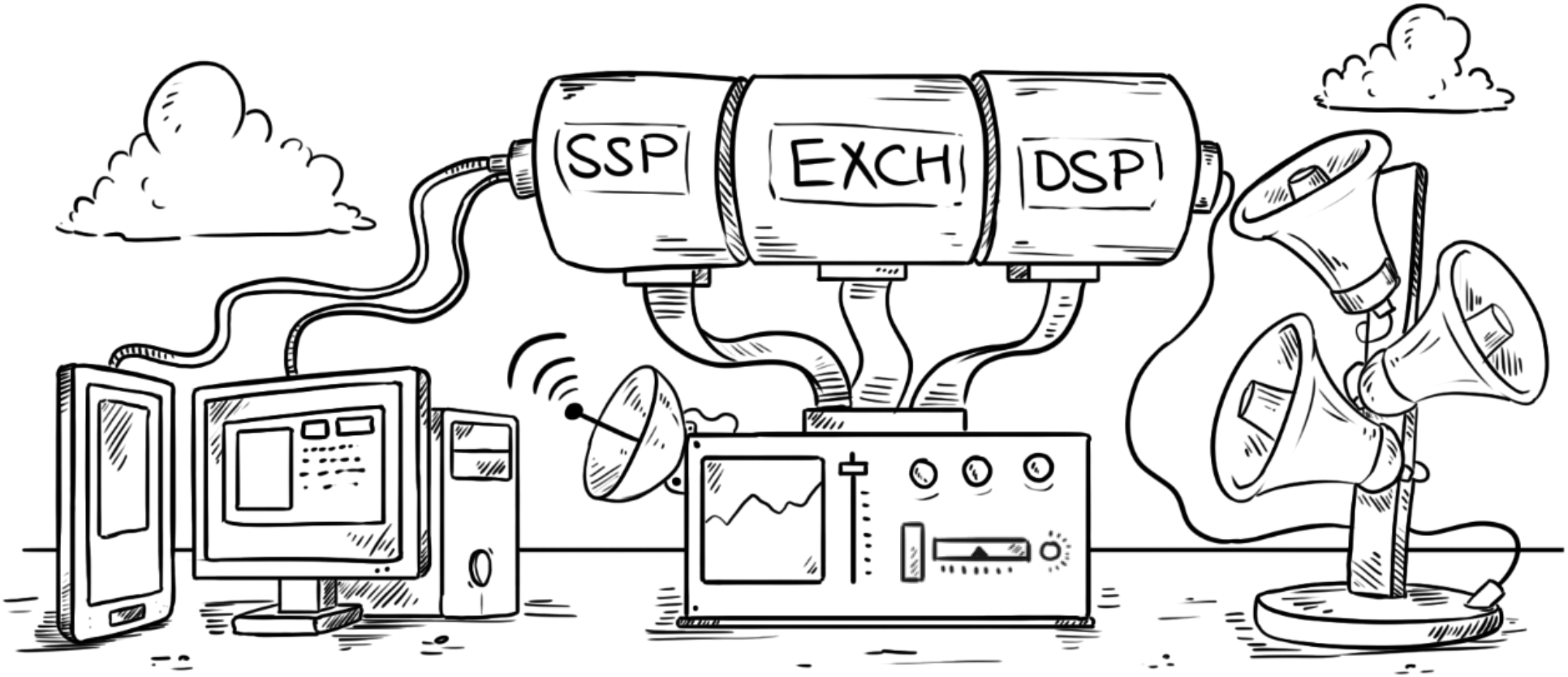
Digital Advertising can be measured precisely

Digital Advertising



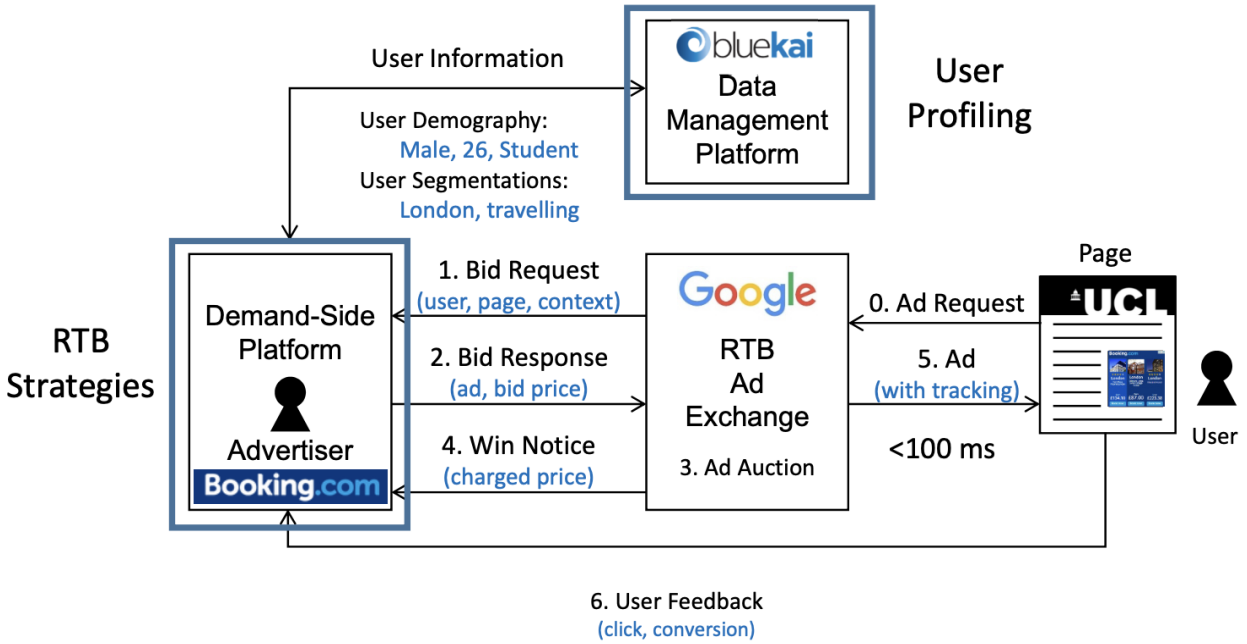
Make the best match between the advertisers and internet users with economic constraints

Digital Advertising Ecosystem



Real Time Bidding

- **RTB is a mechanism for evaluating, buying and selling every online ad view individually and instantaneously.**
- **Using RTB,**
  - **advertisers can spend money to impress specific users in specific contexts**
  - **publishers can sell their traffic to highest offered price**



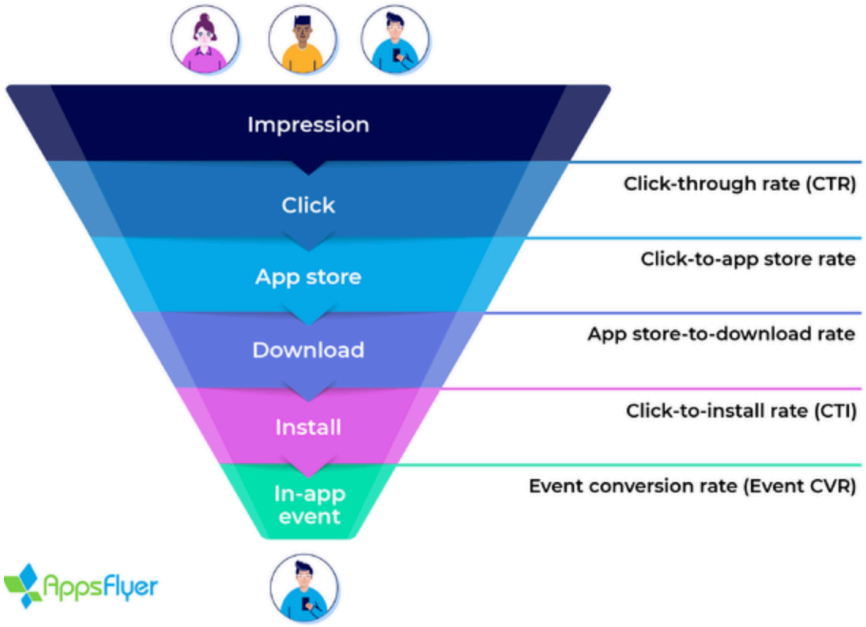
Real Time Bidding

- **Advertisers have different goals with different Pricing Schemas**
  - **CPM: cost per mille impression [favored by publisher]**
  - **CPC: cost per click**
  - **CPA: cost per action [favored by advertiser]**
- **Bidder have to predict the value of each impression for each ad:**

$$E[Rev|a,u] = Pr(click|a,u) * CPC_a$$

Called CVR, **Unknown**! Need to be calculated

Income per Click, **Known**



- **Bidder have to estimate CVR using available data in real time:**
  - **User**
  - **Context**
  - **Ad**

CVR Estimation: Problem Definition

- **Problem Definition**

One instance data

- Date: 20160320
- Hour: 14
- Weekday: 7
- IP: 119.163.222.*
- Region: England
- City: London
- Country: UK
- Search Query: “iphone 6s case”
- OS: Windows
- Browser: Chrome
- Ad title: “iphone 6s case on sale!”
- Ad content: “Customize your case design”
- Bid keywords: “iphone case”
- User occupation: Student
- User tags: Sports

Corresponding label

Click (1) or not (0)?

Predicted CTR (0.15)

- **Challenges:**
  - **High Dimensional Data (order of millions of features)**
  - **Too Sparse Feature Vectors**
  - **Very Imbalanced Classification [The convert events are too rare]**

CVR Estimation: Sample Solution

Information about User and Content

Model:

$$f_\theta(x) \rightarrow y$$

Probability user will click on the content

Simple Logistic Model:

$$f_\theta(x) = \sigma\left(\sum_{k=1}^{d} \theta_k \phi_k(x)\right), \quad \sigma(t) = \frac{1}{1+e^{-t}}$$

Features function extract numeric properties from x

Useful features?

- User Features:
  - age, gender, and click history
- Product Features:
  - Price, popularity, and description…
- Combined (Cross) features:
  - I(20 < age < 30, male, "xbox" in desc)…

e.g., Recurrent NN output..

e.g., Language Model Embedding

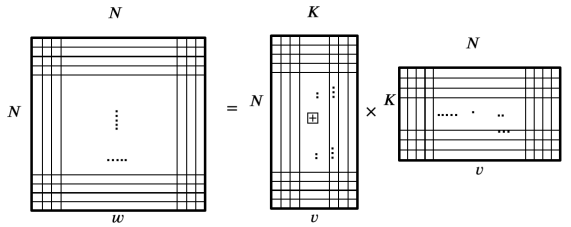
Hand coded features

- **Pros**
  - **Fast Prediction: only one inner Product should be calculated**
  - **Fast Learning Methods: efficient online algorithms**
  - **Interpretable**
- **Cons**
  - **Linear models don't consider correlation among features**
  - **Linear models can only memorize feature combinations with label**

CVR Estimation: Predictive Models

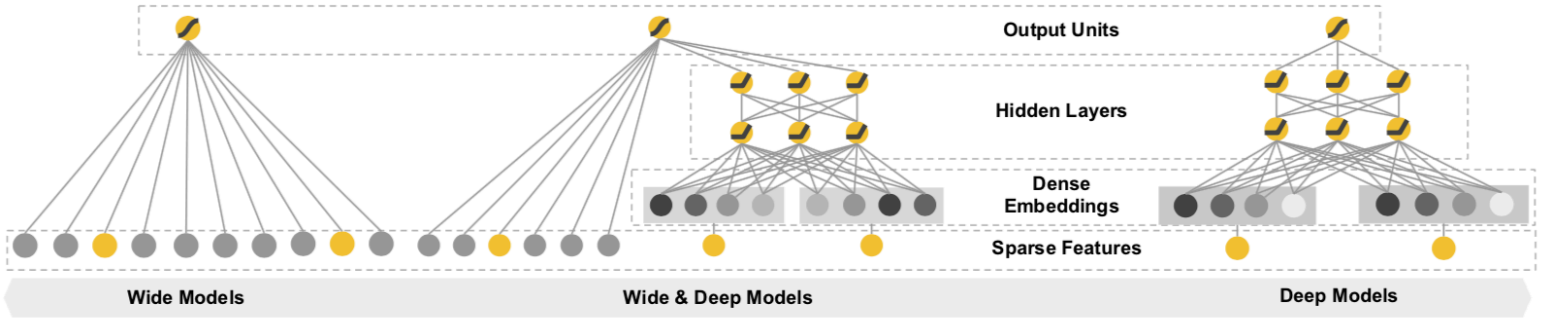
- **Factorization Machines**
  - Fast Prediction
  - Consider Features correlation
  - Learning the parameters is expensive and can't be done online
  - Over-generalization

$$p(y \mid x, w) = f(\phi(x, w))$$
$$\phi(x, w) = \sum_{i,j \in F} w_{ij} x_i x_j$$



- **Deep models**
  - Good generalization and memorization
  - Learning deep models is computationally expensive
  - Time consuming prediction method
    - Deep features need to be calculated in prediction time
    - Can't be scaled to RTB size but can be used in sponsored search

$$P(Y = 1|\mathbf{x}) = \sigma(\mathbf{w}_{wide}^T[\mathbf{x}, \phi(\mathbf{x})] + \mathbf{w}_{deep}^T a^{(l_f)} + b)$$



CVR Estimation Challenges in Production

Data Collection

- Processing a huge amount of noisy feedback
- Data validation and verification
- Preparing collected data for prediction instantly

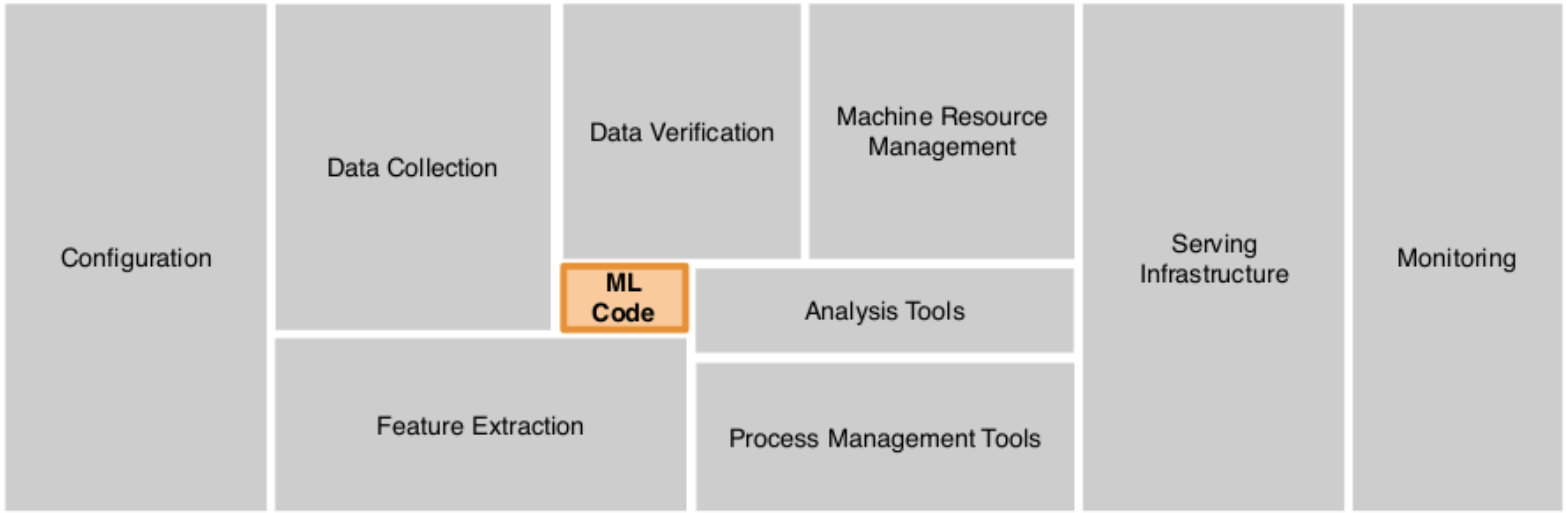
Training

- Data Drift and concept drift
  - continuous training
- Feedback loop in training data
  - Multi-Arm Bandit
- Delayed Feedback for actions

Inference

- Real-time response [<100ms]
- High Throughput [>1B Daily imps]
- Bursty Traffic
  - monitoring, alerting and Auto-Scaling

CVR Estimation System



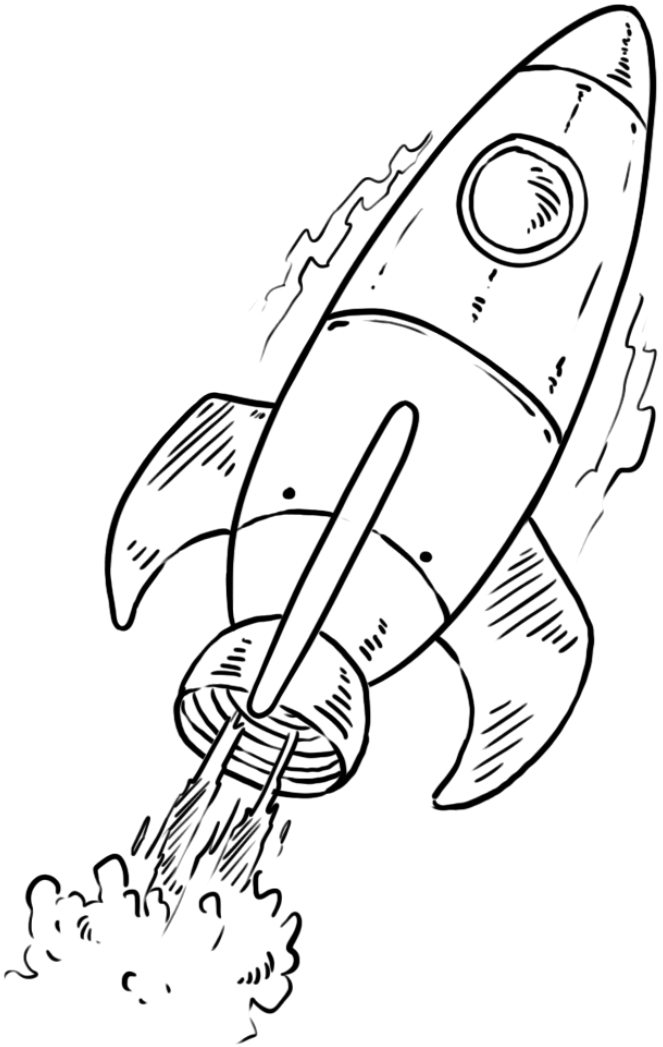
Adopted from Sculley, David, et al. "Hidden technical debt in machine learning systems." *Advances in neural information processing systems* 28 (2015).

Tapsell Brain 1st Generation

- **Business state:**
  - 500K daily impression
  - Video advertising SDK with 50 Publishers
  - CPM and CPC campaigns
- **Technical State:**
  - Centralized system to answer the requests
  - Estimating CTRs using a simple Bayesian Bernoulli Model
  - Visualizing the historical data and improve algorithm incrementally
- **Cons:**
  - Not scalable
  - Large error in CTR estimates
- **Pros:**
  - Best Performance based advertising platform in its own time

Tapsell Brain 2nd Generation

- **Business state:**
  - 1M+ daily impression
  - 150+ Publishers
  - CPI Campaign
- **Technical State:**
  - Adding multi-level cache to response more requests (still centralized)
  - Estimating CVRs in lower granularity
  - Adding time effect to the CVR estimation model
  - Using feedback data to improve CVR estimations
- **Cons:**
  - Not scalable
  - Large error in CVR estimation for post-click actions
- **Pros:**
  - The Only CPI based advertising platform in its own time

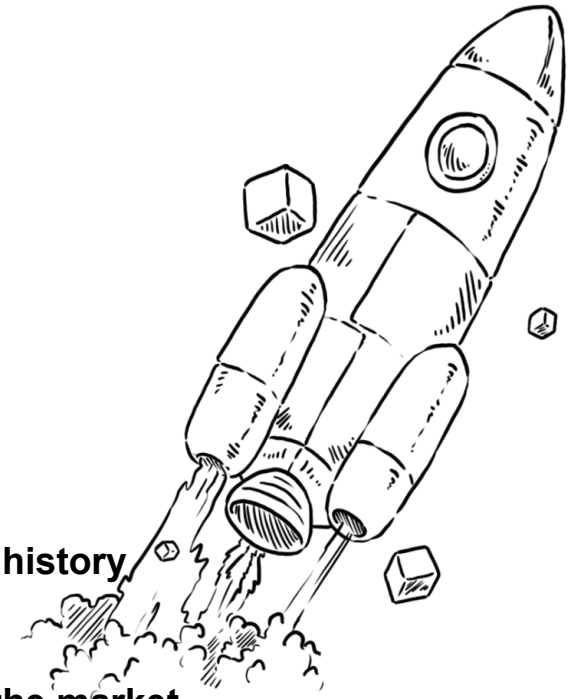


Tapsell Brain 3rd Generation

- **Business state:**
  - 100M+ daily impression
  - 500+ Publishers
  - CPI, CPA Campaign

- **Technical State:**
  - Making the model horizontally scalable in all levels
    - Changing the servers' OS to DCOS
    - Switching to distributed programming platforms (Apache Spark)
    - Switching to distributed Databases (Cassandra, …)
    - Dockerizing all modules
  - Making the CVR estimation model much more efficient by considering all users' history

- **Pros:**
  - The system is completely scalable and there exist no technical limitation to get the market
  - Best Performance based advertising platform in Iran



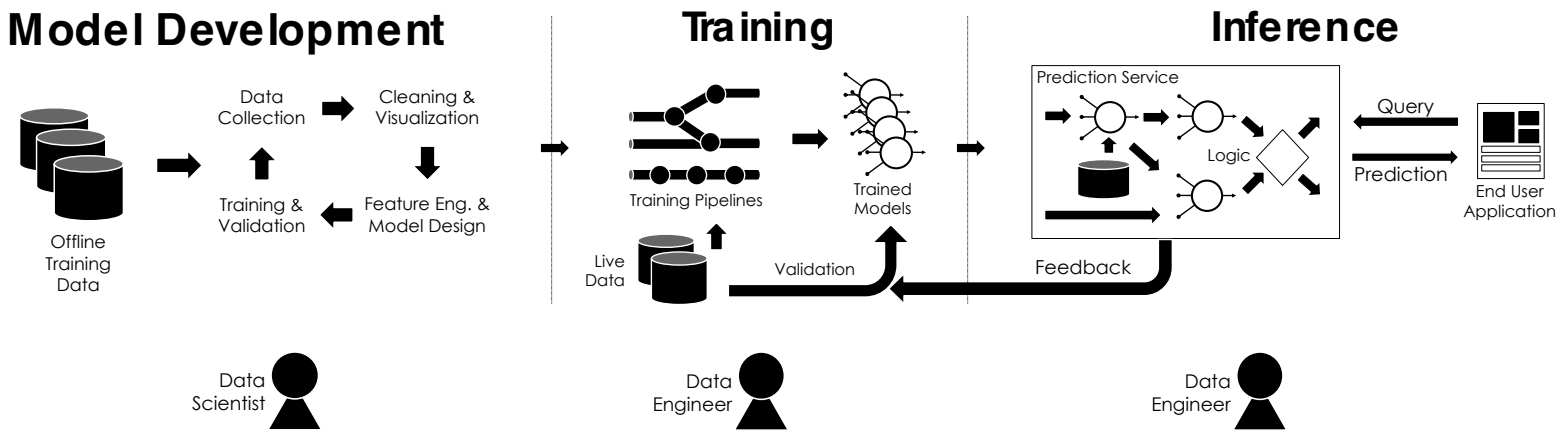
Tapsell Brain 4th Generation

- **Business state:**
  - 1B+ daily impression
  - 20% improvement in ECPM
  - 5000+ Direct Publishers
  - More than 10x traffic in comparison to 3rd generation

- **Technical State:**
  - Model Development automation
  - Training complex models continuously
  - Scalable serving prediction of complex models

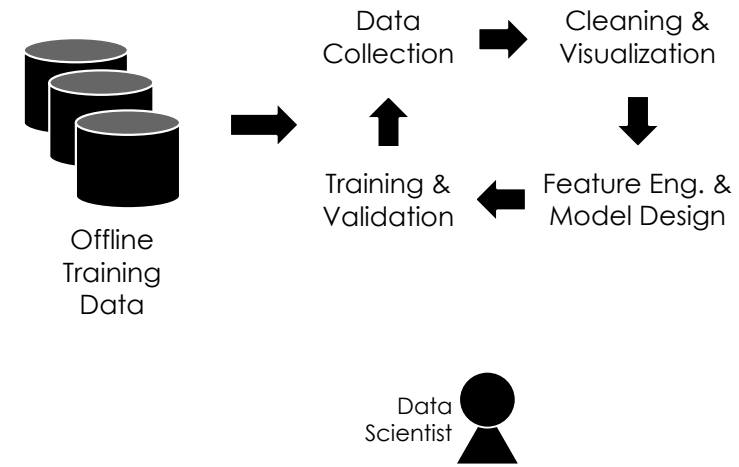
- **Pros:**
  - Be able to improve system by develop, train and serve new models easily

Current Structure & Architecture



Adopted from gonzalez. “AI for Systems and Systems for AI." *Berkeley (2019)*.

Model Development

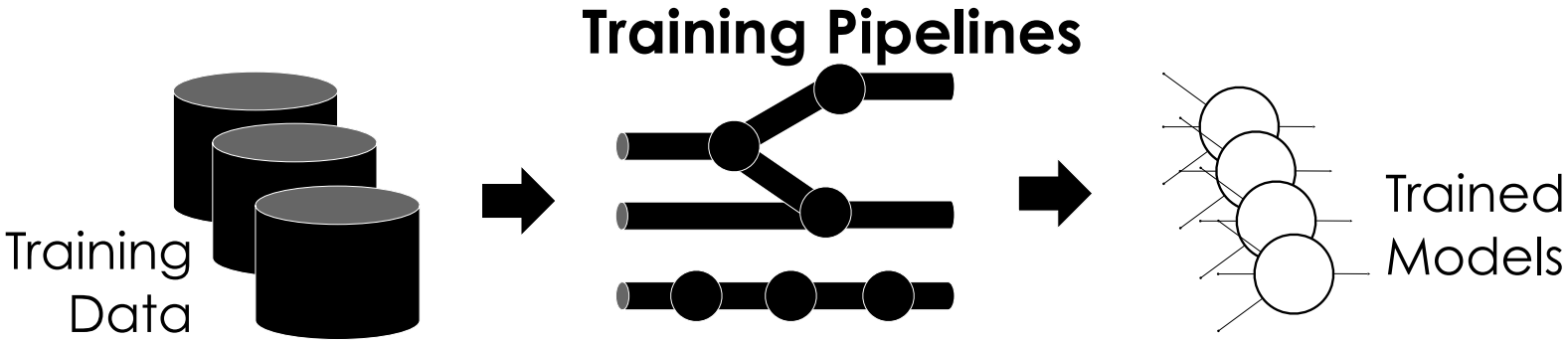


- **Identifying** potential sources of data
- **Joining** data from multiple sources
- Addressing **missing values** and **outliers**
- Data Centric Model Design
- Offline Model Evaluation
  - Designing metrics by considering the bandit feedback
  - Fine-tune hyper-parameters by doing extensive experiments
- Output of model development team are:
  - Dashboards and notebooks (insight)
  - training pipelines

Training Pipeline

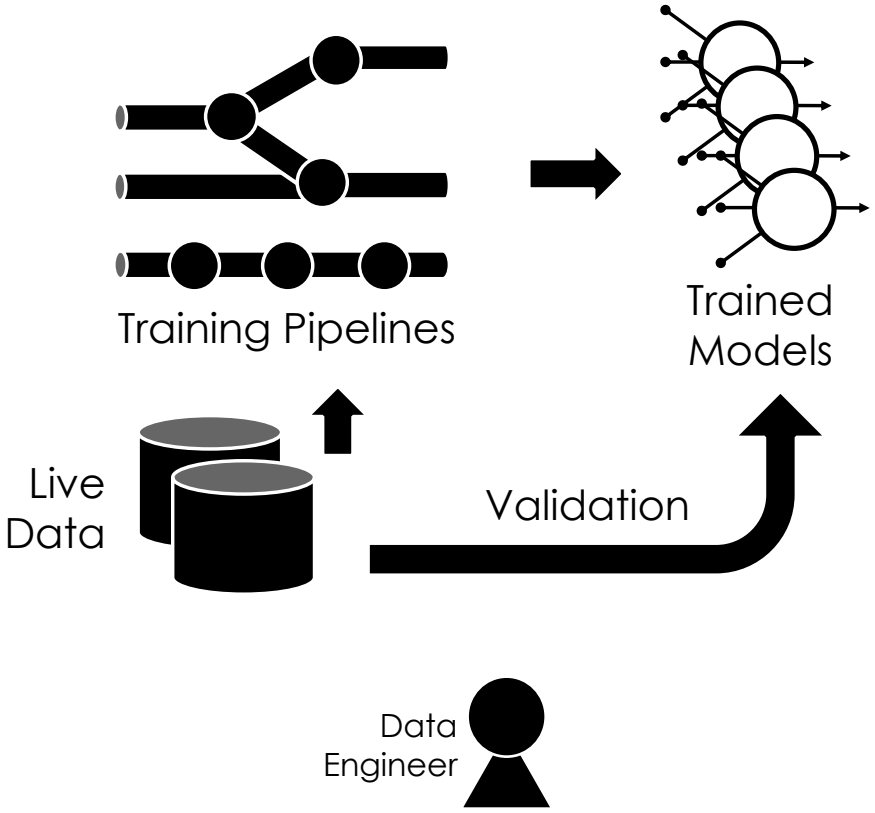
Training Pipelines Capture the **Code** and **Data Dependencies**

- Description of how to train the model from data sources



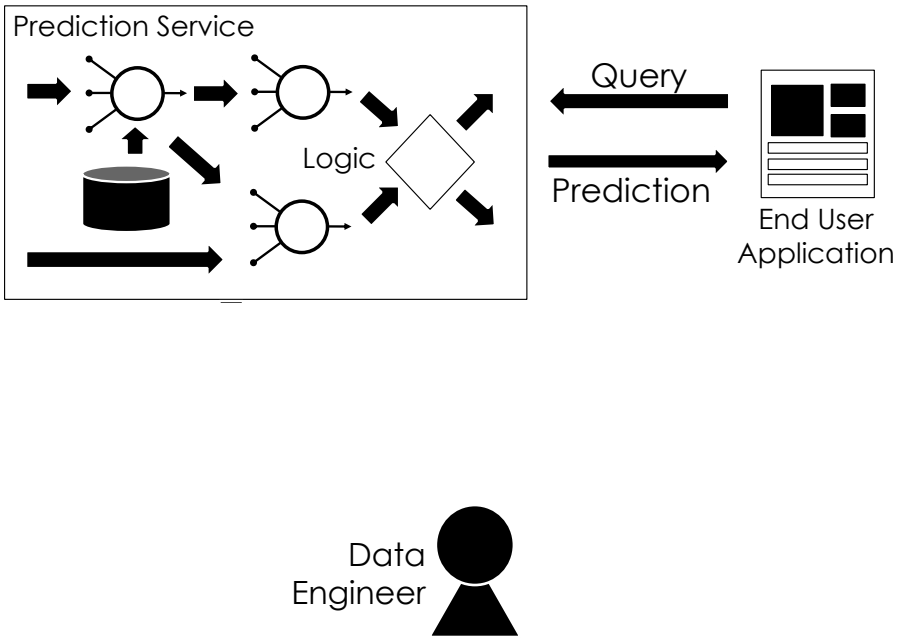
Software Engineering Analogy	Training Pipelines → Code
	Trained Models → Binaries

Training



- Training models **at scale** on **live data**
- **Retraining** on new data
- Automatically **validate** prediction accuracy
- Manage model **versioning**
- Online model evaluation
  - Designing fair A/B tests

Prediction Serving



Goal: make predictions in ~50ms under **bursty** load

- *Complicated* by Deep Neural Networks

Two Approaches

- ***Offline:*** Pre-Materialize Predictions
- ***Online:*** Compute Predictions on the fly

Current Challenges

- Model Improvement
  - We still have a far way to be optimized in CVR estimation
  - Bid estimation in competition to other DSPs is still a new challenge for us

- Model Development and Training Automation
  - Handling huge amount of available features (Feature Store)
  - Distributed training of models (Kubeflow)
  - Train and Serving of Deep Models (TFX)
  - Experiment tracking (ML flow)

How to Join Us

- Research Topic for M.Sc. and Ph.D. students
  - Computational Advertising is a hot topic in top conferences such as KDD, WSDM, WWW, ...
    - Real world problems
    - Real Datasets
    - Baseline Methods that can be used to develop more advanced ones

- Apply for full time or part time job by
    - Send your resume to jobs@tapsell.ir
    - Fill the form at jobs.tapsell.ir

Thank You!