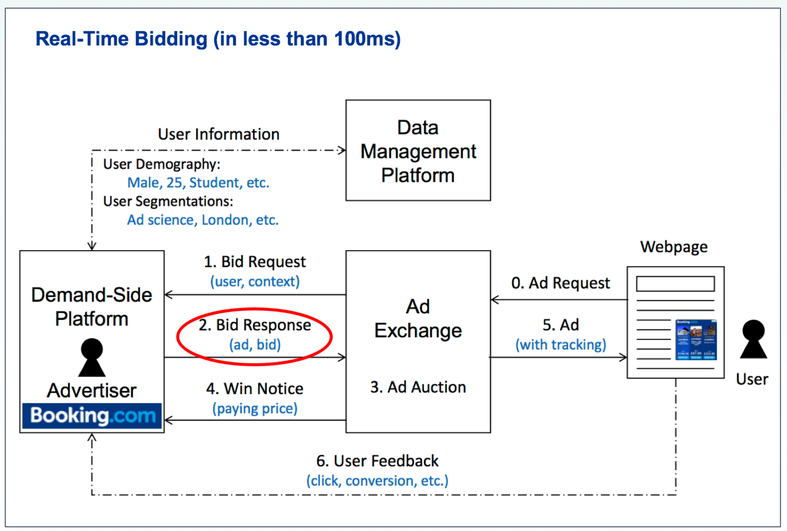
Real-Time Bidding on Advertisements

Internship Information

Name: Anastasia Petrova
Company Name: WalmartLabs
Project: WMX Real-Time Bidding Strategy
Product: Real-Time Bidding Pipeline
Dates: 6/13/16 – 9/20/16
Location: Sunnyvale, CA

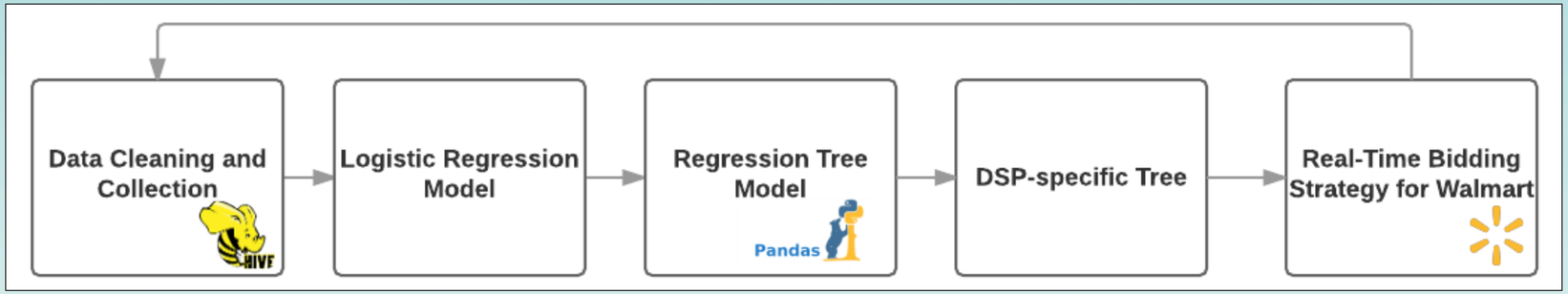
Real-Time Bidding Background

- To reach new audiences, advertisers want to run campaigns for Walmart products on third party sites. This is done via an auction where each advertiser bids on ad space. This project is used to predict how much Walmart should bid on an ad.
- In order to successfully predict how much to bid at an auction, we try to maximize click-through rate (the probability with which a customer clicks on the ad) based on many features.
- My project has been to update the pipeline by which this training and model deployment happens as well as analyze how different components play into performance of the model.



Source: WSDM-2015 tutorial, Jun Wang, Shuai Yuan, and Kaihua Cai

Pipeline



Key Achievements

- Helped deploy multiple advertisement campaigns with our DSP (third-party service that actually places the bids on advertisements), each potentially yielding many new customers. An A/B test was run on each campaign testing our deployed model against the default, which yielded positive results.
- Switched the logistic regression trainer from GLMNet to Vowpal Wabbit, improving training time by 10-fold and accuracy metrics by at least 10% on all test results.
- Improved other parts to make the pipeline more robust in production, including reduced cluster dependencies.
- Ran offline tests to determine optimal feature space to use when training a Logistic Regression model, and experimented with aggregation of data over multiple advertising campaigns.

Relation to UCSD

- Working with new machine learning concepts gave me a clearer understanding of what I would like to study, including a better idea of what classes I want to take and potentially what I would like to continue studying in graduate school.
- Provided me a good background on what real-life applications of concepts I learned in class could be.
- Specific classes that were useful:
  - MATH180A/183: Basic probability and statistics knowledge.
  - CSE151: Introduction to many machine learning concepts.
  - MATH170A: Understanding of numerical linear algebra and gradient descent.
  - CSE132A: Good understanding of databases and SQL (useful when working with HQL).