Notes for LIVE session: Ad Click Prediction

1. For cracking programming, DS and Algo rounds in top product based companies: InterviewPrep Course: https://interviewprep.appliedcourse.com/
2. Real-world problems solving(end to end) in interviews.
3. My experiences at Yahoo! Labs & Amazon.

4. How do we do this session: 2 hour mock interview on Ad-Click prediction for Search.
   a. I will introduce and explain the business context.
   b. I will ask questions and wait for your thoughts/responses on Slack.
   c. 1:n interview/classroom-discussion in which everyone can participate
   d. I will try to simulate an interview as closely as possible.
   e. Let's make it as interactive as possible.
   f. We have seen a ton of code in the course and hence will not prioritize that.

5. Business problem:
   a. E.g: Google Search
   b. Pay per click ads.
   c. Bidding by advertisers: PPC, Search keywords, Ad-creative.
   d. Search engine ranking: Maximize profit (objective) while keeping the results relevant.
   e. Real world constraints: ??
   f. ML problem: predict the probability of click: P(Click| ad, user, context)
   g. Expected Revenue = pClick*bid
   h. Position bias.
   i. Typical CTR: ~4%.

6. Mapping to an ML problem:
   a. What type of problem: classification with probabilistic outputs.
   b. Objective: Min Log-loss
   c. ML constraints: low-latency, interpretability, parallelizable training,
   d. KPI/metric: log-loss, AUC (if you only care about ranking and not pClick values)

7. Dataset:
   a. Search Logs
   b. Format: each row corresponds to an ad impression: <x_i=features, y_i=isCliCk?
   c. How to correct for Position bias?
   d. Features: user, ad, context, historical rates, geo,
   e. Encoding of each feature: categorical features, Numeric features, ….
   f. Dataset properties: imbalanced data: undersampling, reweighting-schemes.
   g. Time-based-splitting, Demographic split,

8. Modeling?
    a. Choice of models: NB, Logistic regression,
    b. NEXT WEEK
        ??
    c. Measurement of model performance: ??
    d. Modeling at Scale: ??

9. Productionisation & monitoring
    a. Choice of architecture: ??
    b. Challenges: ??
    c. Monitoring: ??
    d. Retraining :??
    e. Real-time/Online training: ??
    f. Cold-start:??
    g. State of the art @ Google: https://ai.google/research/pubs/pub41159

10. Real world data and Code:
    a. https://www.kaggle.com/c/kddcup2012-track2
    b. https://drive.google.com/drive/folders/1yRpfqSxojPwmPHVJXSMf-hC8EiOs8hTn?usp=sharing

11. External References:
    a. Google: "ad click prediction"
    b. Google Scholar: "ad click Prediction"
       https://scholar.google.co.in/scholar?hl=en&as_sdt=0%2C5&q=ad+click+predictrion&btnG=
    c. https://towardsdatascience.com/mobile-ads-click-through-rate-ctr-prediction-44fdac40c6ff
    d. https://turi.com/learn/gallery/notebooks/click_through_rate_prediction_intro.html