| vertical | spend % | CTR % | CVR % | yield % | CPM % | CPC % | CPI % |
|---|---|---|---|---|---|---|---|
| vertical 1 | 10.46 | −4.67 | 30.02 | 23.63 | 8.69 | 13.63 | −12.38 |
| vertical 2 | 9.25 | −5.64 | 0.02 | −5.62 | −7.26 | −1.72 | −1.73 |
| vertical 3 | 8.61 | 2.70 | 26.04 | 28 | 6.28 | 3.84 | −17.82 |
| vertical 4 | 9.19 | 2.01 | 2.52 | 4.58 | 4.04 | 1.99 | −0.51 |
| vertical 5 | 20.83 | 1.89 | 21.12 | 23.47 | 9.37 | 11.89 | −7.61 |

**Table 5: Online performance lifts (in %) after adding cross features from MLPs with transfer learning.**

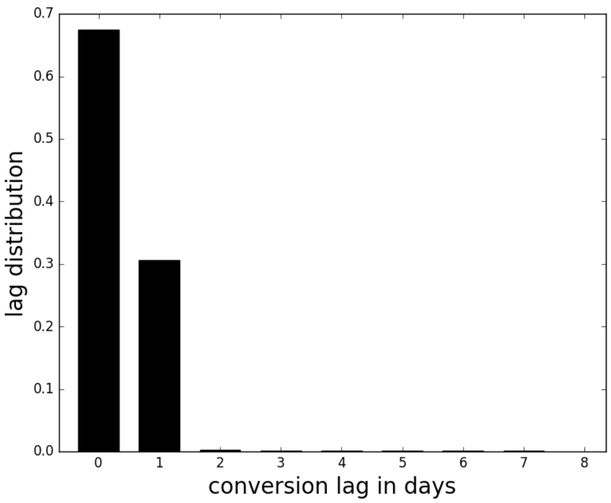


**Figure 4: Conversion lag distribution for app installs (time between click and first use of the app). Majority of the conversions are received within 2 days of the click.**

well during exploration. This was beneficial because advertisers look for high scale in delivery along with low CPIs, and usually expect good results very soon.

## 9 FUTURE WORK

We continue to experiment with more features, modeling improvements, and training. Some of the interesting next steps for us are (a) to have incremental model training with sequential training, (b) optimize training and scoring deep neural networks for a large scale setup like ours, and (c) incorporate long term value of app-installers in our modeling pipeline.

## REFERENCES

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, and et al. Tensorflow: Large-scale machine learning on heterogeneous systems. *Software from tensorflow.org*, 2015.

[2] D. Agarwal, R. Agrawal, R. Khanna, and N. Kota. Estimating rates of rare events with multiple hierarchies through scalable log-linear models. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 213–222. ACM, 2010.

[3] D. Agarwal, A. Z. Broder, D. Chakrabarti, D. Diklic, V. Josifovski, and M. Sayyadian. Estimating rates of rare events at multiple resolutions. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 16–25. ACM, 2007.

[4] A. Z. Broder. Computational advertising. In *SODA*, volume 8, pages 992–992, 2008.

[5] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. In *American Economic Review*, 2007.

[6] eMarketer. US mobile app install ad spending, 2014-2016. *http://www.emarketer.com/Chart/US-Mobile-App-Install-Ad-Spending-2014-2016-billions-change-of-total-mobile-ad-spending/198131*.

[7] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine. ICML 2010.

[8] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1998.

[9] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006.

[10] Y. B. Jason Yosinski, Jeff Clune and H. Lipson. How transferable are features in deep neural networks? NIPS 2014.

[11] N. Kota and D. Agarwal. Temporal multi-hierarchy smoothing for estimating rates of rare events. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1361–1369. ACM, 2011.

[12] H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, S. Chikkerur, D. Liu, M. Wattenberg, A. M. Hrafnkelsson, T. Boulos, and J. Kubica. Ad click prediction: a view from the trenches. KDD 2013.

[13] S. Rendle. Factorization machines. In *IEEE International Conference on Data Mining*, ICDM 2010.

[14] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: Estimating the click-through rate for new ads. WWW 2007.

[15] R. Rosales, H. Cheng, and E. Manavoglu. Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 293–302. ACM, 2012.

[16] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 2015.

[17] Y. Shan, T. R. Hoens, J. Jiao, H. Wang, D. Yu, and J. C. Mao. Deep crossing: Web-scale modeling without manually crafted combinatorial features. KDD 2016.

[18] Statista. Number of apps available in leading app stores as of March 2017. *https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/*.

[19] H. R. Varian. Position auctions. In *International Journal of Industrial Organization*, 2007.

[20] H. Xinran, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, and J. Q. Candela. Practical lessons from predicting clicks on ads at Facebook. *International Workshop on Data Mining for Online Advertising*, 2013.

[21] Yahoo Gemini. *https://gemini.yahoo.com/advertiser/home*.

[22] Yahoo Gemini. Drive app installs on mobile. *https://developer.yahoo.com/gemini/advertiser/guide/adcreation/drive-app-installs-mobile/*.