



CASE STUDY



IS IT POSSIBLE TO PREDICT ADVERTISING SUCCESS?

Company Info

Name: Avito

Location: Moscow, Russia

Industry: Advertising

When an employee at any company starts work, they first need to obtain the computer access necessary to fulfill their role.

Placing an ad is not easy, but we all know it is critical when selling something. Before shifting to online sites, classifieds were located according to different categories. With the Internet and the availability of large amounts of data, there are many more factors to consider.

Technologies have also evolved, and now there are AI tools capable of profiting from the data available. These facts moved Avito, one of the leading Russian classified sites, to search for a solution that could efficiently predict if an ad would be clicked or not. With an average of over 70 million users per month, Avito has a big task: to connect buyers and sellers from all corners of the biggest country on earth. It provides its service by having a web site with sections for general goods for sale, jobs, real estate, personals, cars for sale, and services. There are three different types of ads available: regular, highlighted, and context. Context ads are organized according to statistical factors. However, this method ignores the behavior of different users and, thus, is not optimal.

Avito wanted to change this and prepared a set of data to create a model capable of predicting if an individual user would click on a specific ad.

"We wanted to find an efficient solution, in a short time."



LogicPlum Clicks In

LogicPlum's platform brings three essential components to AI: speed, efficiency, and cost-effectiveness. It does this by using cutting-edge algorithms, which are updated continuously, by being capable of automating model creation; and using a cluster of open-source technologies.

Complementing its platform, LogicPlum has a team of data scientists capable of tackling the most challenging problems. It was this team that realized that this problem could be solved by using LogicPlum's platform. As one of them said: *"We established a tough goal for ourselves. We wanted to find an efficient solution, in a short time"*.

Organizing the Data

"The first thing that we noticed was that we had to act systematically," explained the team lead. "We had to begin by analyzing the data available."

A quick review of the data revealed that its distribution was unbalanced. *"Therefore, we used a downsizing technique to correct this problem,"* added the data scientist. *"When you have an unbalanced sample, the model will tend to learn more from the abundant data, thus, creating a potentially biased solution."*

The dataset contained information about users and their behavior, such as the device and the search parameters used, whether the user had clicked on an ad or not, and data about its geographical location. Besides, it had information about the ads, such as title,

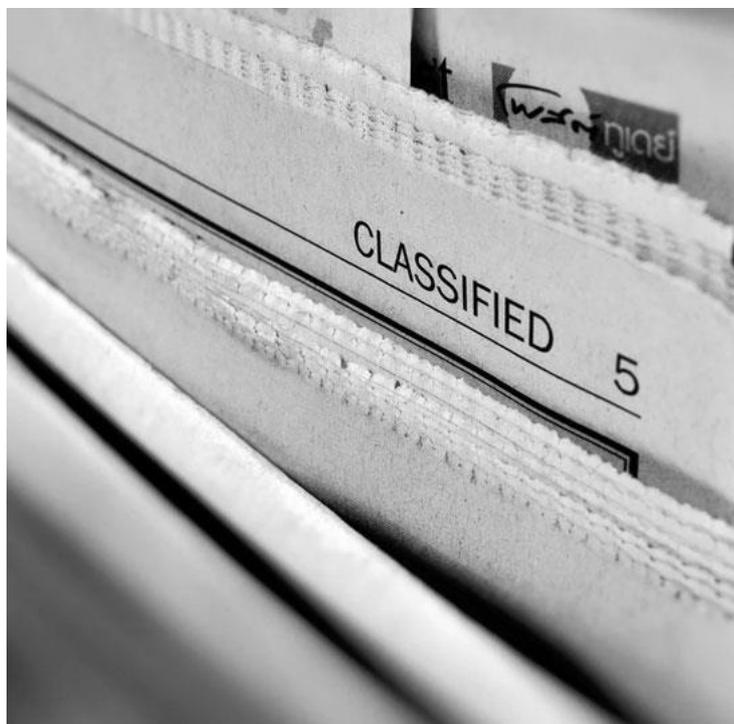
location in the web page, category, subcategory, whether it was a context ad or not, and the item's price. "We created a database with the organized data," said the team lead, *"because we had to consider the vast amount of data."*

Finding the Right Solution

"We did some literature investigation because this was a new problem for us," added the team lead. *"We found that many scholars suggested the use of Field Aware Factorization Machines or FFM, because they are capable of handling large, sparse sets of data containing several categorical fields. And, we thought to start from there"*.

The team let LogicPlum's platform do the hard part: finding a collection of models and ordering them according to their prediction accuracy. Once they had these hundreds of algorithms, they generated a final model by combining several models predictions via a neural network.

The strategy proved to be successful, as they saw that





the end-result could predict if a user would click on a specific add with a 0.04 LogLoss when given the necessary information about the user and the ad. A perfect LogLoss score is 0.00.

☆ Beginning the AI Journey

The team wanted to create a solution that would allow the user to manage the complete AI journey. They began by preparing the data to be delivered so the user could re-use, reorganize, or modify it. As LogicPlum accepts inputs from many sources such as Hadoop, the most popular databases, and CSV files, they decided to leverage the platforms ODBC connections for their database.

Following this, LogicPlum created a blueprint that included all the necessary information to explain how the platform had arrived at the final solution. This *"was very important, as the logic behind the solution can provide important clues for data-driven decision making. Besides, it showed that overfitting was not the reason for such a good prediction score,"* said the team lead.

Finally, it was necessary to provide a way for a user to connect to the platform. This was achieved via the single-point API that LogicPlum provides to its clients. *"This allows for the use of the platform by a user who doesn't need to be an AI expert, providing cost-savings and easy accessibility,"* explained one team member.



Contact Us

LogicPlum
1550 West
McEwen Drive
Suite 300
Franklin, TN 37067USA

www.logicplum.com
message@logicplum.com

LogicPlum

©2020 LogicPlum, Inc. All rights reserved. LogicPlum and the LogicPlum logo are trademarks of LogicPlum, Inc. All other marks are trademarks or registered trademarks of their respective holders. Avito did not participate in this case study. LogicPlum does not represent this case study as an endorsement by Avito of LogicPlum or any of LogicPlum's services.