AdCypher

Understanding digital creative performance using machine learning

Ian Kurtz

Project Description:

With digital advertising spend surpassing one hundred billion dollars annually, the industry's need for increased measurement and insights particularly in the creative space has become critical. When designing a campaign teams build strategies to alter consumer perception of a brand and design the verbiage, calls to action, or the color schemes and layout to echo the message a brand. Understanding which components within the creative are working, is a difficult challenge with very few descriptive products in market. AdCypher uses machine learning to scan thousands of images to identify whether a creative will be successful and to provide some insights as to what is driving the performance.
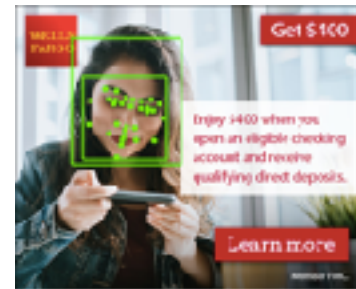
Data Pipeline:

AdCypher scrapes two websites and connects into two APIs to build out its data pipeline. The first step is gathering information regarding a wide variety of entities including businesses, colleges, and governments agencies, who conduct marketing campaigns in the US. Information about the age of the entity, size, web traffic, industry and more are scraped. Next, the display advertisements from those companies from 2014 to present are scraped and stored locally. Finally, those images are scanned to understand the spacing, color scheme, verbiage, and content used in the ad. Finally the keywords used in each ad are pushed into Google trends to see if consumers searched for those keywords while a campaign was on versus when it was off. Campaigns that successfully drive lift in Google Trends are labeled as successful and used as targets in the model.

crunchbase ➡️ MOAT ➡️ Google AI ➡️ Google Trends

Modeling:

Multiple classification algorithms including Random Forest, Logistic Regression and Gradient Decent were run over the data to see if the model's features were able to predict if a campaign will be successful. Topic modeling techniques such as SVD were used to interpret the text used in the ads. After tuning hyper parameters the Random Forest model generated an accuracy score of 81%. Verbiage, age of company, campaign length, and spacing were the most important features in determining a campaign's performance.

Interesting Findings:

- The most commonly used color, blue, actually decreases the likelihood of an ad being successful by 6%.
- Educational terminology such as the word learn leads to a greater chance of consumers googling the company.
- Including a photo of a consumer smiling and engaging with the product remains one of the most powerful contributors to a successful campaign
- Campaigns which last between one and four months are top performers.
- Companies who are between 0-10 or 20-30 years old generate the highest increases in Google searches

Ian Kurtz
Data Scientist

San Francisco, CA
510-666-5556
IanKurtz@gmail.com

DATA SCIENCE PROJECTS

• Image classifcation using Google AI to predict the performance of a digital marketing creative. Project employed multiple web crawlers to capature images and information about the companies delivering advertisements.
• Sale price prediction of heavy equipment at auction based on its usage, equipment type, and configuration
• Churn prediction with a ride-sharing company using Logistic Regression, Random-Forest Gradient boosting models
• Classification of spam emails using Natural Language Processing TF-IDF vectorization and Naive Bayes classifier
• Fraud Detection System using Tensor-flow Deep Neural Network on AWS Server

RELEVANT EXPERIENCE

Managing Analyst, Uber
San Francisco, CA
AKQA
Feb18–Jan 19
•Designed measurement stack for Uber's digital brand campaigns
•Managed a team of three analysts including hiring, training, and delivery management
•Built a user campaign frequency model using SQL to optimize performance
•Consulted on design for Uber's brand marketing data layer

Lead Programmatic Analyst
San Francisco, CA
AKQA
Jun17–Jan 18
•Audited on-target percentages for over sixty second and third-party data vendors using comScore's vCE, Nielsens DAR, and Survata surveys which resulted in a 20% increase in accuracy
•Aggregated and analyzed two years of video performance data across multiple channels to enhance client strategy recommendations
•Ran multiple MTA studies using Shapley methodologies to enhance fiscal year planning recommendations

SRM Analyst, Clorox
San Francisco, CA
AKQA
Feb 16–Jun 17
•Manage the planning, design, and development of client A/B and Multivariate (MVT) tests including detailed test plans and test reports
•Designed multiple BI dashboards and media performance decks to enhance campaign learnings•Operationalized both brand engagement and FOAS campaign targeting strategy for multiple clients
•Worked closely with media, programmatic, and search teams to proactively strategize learning agendas, optimization campaigns, and budget allocations
•Vetted and on-boarded analytics tools including Survata SF DMP, Kartar, and Nielsen MTA

Growth Analyst
San Francisco, CA
WRKSHP
Feb 15 –Jul 15
•Increased return on aggregate spend (ROAS) by 12% managing seven mobile ad networks
•Developed and managed six analytic dashboards focused on cohort analysis, optimizing ROAS, monitoring eCVR performance, payer rates, and paid vs organic install ratios

Business and Data Analyst
San Francisco, CA
IPG Mediabrands
May14–Sep 14
•Aggregated and harmonized 180 million records across 17 markets into a Netezza SQL server to help management identify international shifts in consumer media consumption
•Designed and published Tableau dashboards built to optimize executive decision making

SKILLS

Programming
• Python
• Pandas
• SQL
• Git
• Numpy
• Scikit-learn
• Surprise

Technical
• Tableau
• Machine Learning
• A/B testing
• Statistical analysis
• Data analysis
• Web scraping

Soft
• Business acumen
• Oral and written comm
• Strategic planning

Education

Galvanize Data Science Immersive

University of California, Berkeley
Bachelor of Science
Business Admiration (GPA: 3.78)