

Predicting Sales Conversion Rates through Collective Behavior and Customer Demographics: A Machine Learning Approach

Zachary Picard

Advisor: Alfa Heryudono

DSC 498 Capstone Project, University of Massachusetts Dartmouth, MA, USA

ABSTRACT:

Conversion rates measure how successful digital marketing efforts are at converting visitors into customers, making them an essential metric to understand and optimize (1). Introducing Machine Learning into this mix revolutionizes Conversion Rate Optimization (CRO) efforts, offering both valuable insights and precision (2). Certain recorded metrics, such as bounce rate and click-through rate, are key to predicting conversion rates (2). The calculated conversion rates can be leveraged to target existing audiences more effectively and engage new audiences through tailored marketing campaigns. This leads me to a few questions I want to explore: Can we use ML models to predict which customers are most likely to purchase specific items using behavioral and demographic data? Can I create a relational database to better train these models? What data is most critical to our predictive model? If we can accurately predict which customers are likely to purchase a specific product, we can extend that insight to identify other customers who may also be interested in buying the same product. Additionally, for customers with lower predicted conversion rates, we can recommend alternative products that are more likely to meet their preferences.

In this project, I will leverage publicly available datasets, SQL databases, and Python to predict customer behavior and maximize profit. The project will focus on calculating the conversion rates of specific products for different customer segments. This information will then be used to optimize ad campaign audiences, improving targeting and maximizing the effectiveness of marketing efforts. Specific ML models I plan to explore and use include: Logistic Regression, Decision Trees, Random Forest, and K-Nearest Neighbors (KNN). I also plan to investigate the efficiency of various algorithms, explore the impact of using multiple models simultaneously, and analyze the discrepancies that arise from overfitting. This will help in understanding how different approaches perform and how to balance model complexity with predictive accuracy.

MAJOR PROPOSED MILESTONES:

- 27 September: Find sourceable, sufficiently large datasets, and devise a plan to create a relational database.

- 11 October: Implement the database strategy and conduct basic exploratory data analysis (EDA) on the dataset.
 - 18 October: Research and learn about ML models and packages: Logistic Regression, Decision Trees, Random Forest, K-Nearest Neighbors (KNN).
 - 22 November: Create and train the models, and use multiple models simultaneously.
 - 6 December: Evaluate the accuracy of the models and uncover discrepancies arising from overfitting.
-

REFERENCES:

- [1] Robson, B. "Boost Conversion Rate Optimization Delivered with Machine Learning." *Web Design Malaysia | WDD Stands for Work Done Differently*, 21 July 2024, wdd.my/blog/conversion-rate-optimization-with-machine-learning/. Accessed 16 Sept. 2024.
- [2] "What Is a Conversion Rate? Definition, CRO, and Importance." *Amplitude*, amplitude.com/glossary/terms/conversion-rate. Accessed 16 Sept. 2024