Chapter 16
A Case Study to Improve Data Vendor Selection

Rick McGraw
Black Oak Partners, LLC, USA

ABSTRACT

All financial services companies use external sources of consumer income data to assess credit risk, develop risk mitigation strategies, and create pre-approved marketing offers. International credit card issuers spend more than a billion dollars in marketing and invest hundreds of millions in product development based on data from third party vendors. The accuracy of income data is the foundation for their business decisions. The purpose of this engagement was to evaluate the accuracy of third party data providers by measuring the data across six dimensions – Accuracy, Relevancy, Completeness, Coverage, Effectiveness, and Cost. The data provided by the three vendors for this project was incorrect by more than 20% on 75% of the records. The goal of the project was to demonstrate the ability to improve the accuracy of multi-source income data used in credit card marketing applications.

THE CHALLENGE

- Data has become a commodity – broad availability, similar characteristics;
- Generally, commodities carry metrics-based standard grades to gauge suitability, but today, data does not;
- Without an established baseline and score, it is impossible to calculate the quality, and therefore, the unique value and contribution of a specific data source.

The inability to resolve data conflict leads to less effective models and sub-optimal decisions.

BACKGROUND

Black Oak analyzed three data vendors under three different operating scenarios and measured several performance metrics for each. The analysis quantified the value each data source provided to any user-defined, decision-making process.

We will refer to the data sources used in the analysis as 'Source A,' 'Source B' and 'Source C'. Each vendor was presented with the same set of records for income append. The results were compared to a known truth file created by Black Oak and its' client. The vendor results were compared to the truth file to determine accuracy. A tolerance range was allowed at 10% and 20% to the vendors (see Relevancy). Vendor records that were within the tolerance interval of the true result were considered accurate.

The operating scenarios were Mortgage Applications, Credit Card Applications and Employee Records. The Employee Records were used as a control on the measurements since the income for the client's employees was known. This data was used to control for potential variances present in the Mortgage and Credit Card Applications.

PERFORMANCE MEASURES

Performance measures were computed for each vendor under each operating scenario. The performance measures computed were the six dimensions of Accuracy, Relevancy, Completeness, Coverage, Effectiveness and Cost. The sections below detail the performance measures and data analyzed.

Completeness

Completeness measures the number of records appended by a vendor that have income information For this particular test, only one element was provided by the vendor –income, so completeness percentages were very high. However, in most cases, vendors are asked to provide multiple elements and they can be measured per element or at a higher composite level.

Coverage

Coverage measures the number of records with income information appended by a vendor to the total number of records requested. This is a direct measure of the utility of the vendor data. A vendor with low coverage may be very good data, but with very few records.

Accuracy

Accuracy measures how close the reported income is relative to the true value.

Two of the data vendors provided an income range rather than a specific value for income. Records are considered accurate if the true value is within the income band provided by the vendor.

One vendor did not provide an income band, but instead provided a specific value. For this vendor, each income was matched to an equivalent band from the other vendors and the result is treated as if the vendor had reported the band.

Relevancy

Accuracy measured the vendor performance against the bands set up by the vendor. Because the vendor sets up the bands, there is a subjective component to accuracy. For example, if one vendor reports many small bands while a second vendor reports fewer larger bands, the second vendor may be measured as more accurate. This is a misleading result because the first vendor reports more precise results and this precision is typically desired.

Relevancy removes the vendor component and measures performance against a tolerance range for truth. For vendors reporting income bands, the result is considered relevant if the midpoint of the band is within the tolerance level of the true value.

Related Content

The Adoption and Utilization of Electronic Business in Response to the Global Economy During COVID-19

Joseph Bamidele Awotunde, Roseline Oluwaseun Ogundokun, Emmanuel Abidemi Adeniyi, Sanjay Misraand Gbemisola Janet Ajamu (2022). *International Journal of Business Analytics (pp. 1-20).*

www.irma-international.org/article/the-adoption-and-utilization-of-electronic-business-in-response-to-the-global-economy-during-covid-19/288518

The Role of Information and Communication Technology in Competitive Intelligence

Dirk Vriens (2004). *Information and Communications Technology for Competitive Intelligence (pp. 1-33).*

www.irma-international.org/chapter/role-information-communication-technology-competitive/22559

Conceptualizing the Role of Data Analytics and Technology in E-Governance: An Insight

Parag Sunil Shuklaand Mayank Mathur (2020). *International Journal of Business Analytics (pp. 1-12).*

www.irma-international.org/article/conceptualizing-the-role-of-data-analytics-and-technology-in-e-governance/246025

Discovering Data and Information Quality Research Insights Gained through Latent Semantic Analysis

Roger Blakeand Ganesan Shankaranarayanan (2012). *International Journal of Business Intelligence Research (pp. 1-16).*

www.irma-international.org/article/discovering-data-information-quality-research/62019

Business Intelligence for Human Capital Management

Maria José Sousaand Ivo Dias (2020). *International Journal of Business Intelligence Research (pp. 38-49).*

www.irma-international.org/article/business-intelligence-for-human-capital-management/245661