

# MISQ Archivist

## Using Forum and Search Data for Sales Prediction of High-Involvement Products

*Tomer Geva, Gal Oestreicher-Singer, Niv Efron, and Yair Shimshoni*

---

### Abstract

A large body of research uses data from social media websites to predict offline economic outcomes such as sales. However, recent research also points out that such data may be subject to various limitations and biases that may hurt predictive accuracy. At the same time, a growing body of research shows that a new source of online information—search engine logs—has the potential to predict offline outcomes. We study the relationship between these two important data sources in the context of sales predictions. Focusing on the automotive industry—a classic example of a domain of high-involvement products—we use Google’s comprehensive index of Internet discussion forums, in addition to Google search trend data. We find that adding search trend data to models based on the more commonly used social media data significantly improves predictive accuracy. We also find that predictive models based on inexpensive search trend data provide predictive accuracy that is comparable, at least, to that of social media data-based predictive models. Last, we show that the improvement in accuracy is considerably larger for “value” car brands, while for “premium” car brands the improvement obtained is more moderate.

**Keywords:** Search trends, forums, social media, word-of-mouth, consumers’ interest, sales prediction, online data, high involvement products