Will Your Event Sell Out? Answering a Prediction Problem Using Machine Learning Algorithms

ECON 483: Economic Applications of Machine Learning

Daniel Bathaei Amanda Muirhead

August 15, 2023

While by itself it is a piece of worthless paper, a ticket grants individuals access to an event that is otherwise closed to the public. If an event is well-liked, many people will buy tickets for it. If a performer, sports team, or exhibition achieves fame and popularity, more people might want to attend their events than are slots available. Once every ticket for an event has been purchased, that event is sold out. To performers, selling out is a sign of popularity. To producers, it is a sign of a missed opportunity. They could have charged more for their tickets, booked a bigger venue, or offered VIP packages. The ability to accurately anticipate whether an event will hit full capacity can have significant implications on a business' demand forecasting, market analysis, and revenue optimization, not to mention the subsequent effects on consumer spending and labour demand. Consequently, producers are asking whether a set of variables such as price and location can predict if their event will sell out. To this end, we use machine learning models to predict whether a show will sell out.

The onset of machine learning in economics, with its capacity to discern patterns from intricate data, has become an indispensable tool in tackling complex prediction problems. Though it has several varying definitions, the study of machine learning boils down to computer programs' abilities to learn and adapt to new data without the interference of humans (**?**). To predict the likelihood of an event selling out of tickets, we will employ several machine learning techniques, namely Classification Trees, Logistic Regression, Random Forests, Support Vector Machines, and Neural Networks.

Background and Current Literature

Event tickets have been in the news a lot lately, namely for the escalating rates at which they are being priced. Event-goers around the world are increasingly lamenting over the rising costs of event tickets, as reports come out of potential price gouging and fees as high as 78% of the original listed price (**?**). While in 2018 it was reported that ticket sales had doubled since the 1990's (**?**), the percentage increase is now estimated to be near triple their pre-Y2K levels (**?**), proving that forecasting the demand for events is already one of many lucrative strategies being employed by artists and ticket-distributing giants such as Ticketmaster. Currently, event tickets sales are a $78bn USD industry, expected to show a growth of 9.7% by 2024 (**?**).

To contextualize our research, we studied existing literature surrounding event sell-out and other related quantity demanded predictions. Previous studies have illuminated various factors that contribute to event success, encompassing variables such as ticket pricing, marketing strategies, historical attendance data, and socioeconomic indicators (**?**).

Limitations

While writing this paper, we observed that companies work very hard to conceal any information related to sales volume and quantity of tickets demanded. StubHub has stopped providing businesses and researchers with access to its application programming interface (API) and only provides API services for producers for their own posted tickets[1] Ticketmaster has recently removed any indicator of sales volume on their API database. Companies like Pollstar[2] charge considerable fees before providing any information on ticket sales. This further solidifies the current importance and relevance of ticket sale information.

Methodology

We use publicly available resources[3] to prepare the data for our models. Pursuant to our objective, we collected data from SeatGeek using the website's application programming interface (API). Similar to other websites, SeatGeek has reduced their available sales volume data. However, SeatGeek still provides some insight on sales volume. Although the company has announced their plan to move away from disclosing any sales volume information through their API, this plan is yet to be implemented.

SeatGeek

SeatGeek currently provides information regarding the volume of tickets sold by an event, performer, and/or venue. This information is shared through the 'score' variable. The 'score' variable is a scaled[4] measure of an event, performer, or venue's ticket sale volume relative to their type. This means that if an event sells more tickets than any other event similar to its type, that event will receive a score of 1. If an event makes no sales, that event will have a score of 0. For each observation we can gather an 'event score', 'venue score', and a performer score for each performer in that event. This information is highly valuable since it is derived from the quantity of units sold. Conveniently, SeatGeek analyzes the social media following of each event's performers, and provides a 'popularity' value based on how popular an events performers are. This value is also scaled between 0 and 1. Consequently, the variables we gather from SeatGeek are shown on Table 1:

[1] https://developer.stubhub.com/api-reference/sales#tag/Sales
[2] https://www.pollstar.com/subscribe)
[3] Ethical obligations dictate that we only use publicly available data, even if we use scraping methods.
[4] $0 \leq score \leq 1$