Cover Page

Lei Wang (Corresponding Author)
Assistant Professor, Smeal College of Business, Pennsylvania State University
426 Buisness Building, University Park, Pennsylvania 16802
Email: luw21@smeal.psu.edu
Phone: 814-867-5838

Ram Gopal
Pro-Dean for Research, Engagement, and Impact
Professor of Information Systems and Management
Turing Fellow, The Alan Turing Institute
Warwick Business School, University of Warwick
Email: ram.gopal@wbs.ac.uk
Phone: 024-761-50164

Ramesh Shankar
Associate Professor, School of Business
University of Connecticut, Room 369, Unit 1041
2100 Hillside Road, Storrs, CT 06269
Email: Ramesh.shankar@uconn.edu
Phone: 860-486-5217

Joseph Pancras
Associate Professor, School of Business
University of Connecticut, Room 328, Unit 1041
2100 Hillside Road, Storrs, CT 06269
Email: joseph.pancras@uconn.edu
Phone: 860-486-0810

Forecasting Venue Popularity on Location-Based Services Using Interpretable Machine Learning

Abstract

Customers are increasingly utilizing location-based services via mobile devices to engage with retail establishments. The focus of this paper is to identify factors that help to drive venue popularity revealed by location-based services, which then better facilitate companies' operational decisions, such as procurement and staff scheduling. Using data collected from Foursquare and Yelp, we build, evaluate, and compare a wide variety of machine learning methods including deep learning models with varying characteristics and degrees of sophistication. First, we find that support vector regression is the best-performing model compared to other complex predictive algorithms. Second, we apply SHAP (Shapley Additive exPlanations) to quantify the contribution from each business feature at both the global and local levels. The global interpretability results show that customer loyalty, the agglomeration effect, and the word-of-mouth effect are the top three drivers for venue popularity. Furthermore, the local interpretability analysis reveals that the contributions of business features vary, both quantitatively and directionally. Our findings are robust with respect to different popularity measures, training and testing periods, and prediction horizons. These findings extend our knowledge of location-based services by demonstrating their potential to play a prominent role in attracting consumer engagement and boosting venue popularity. Managers can make better operational decisions such as procurement and staff scheduling based on these more accurate venue popularity prediction methods. Furthermore, this study also highlights the importance of model interpretability which enhances the ability of managers to more effectively utilize machine learning models for effective decision making.

Keywords: Location-Based Services; User Engagement; Venue Popularity Prediction; Interpretable Machine Learning; SHAP Value;

Received: July 2021; accepted: March 2022 by Subodha Kumar after three revisions.

1. Introduction

Location-based services (LBS) are information services that are accessible with mobile devices and utilize the ability to make use of the location of the mobile devices (Virrantaus et al. 2002). LBS transforms how businesses interact with consumers and provides an emerging channel to reach out to their customers. Besides leaving comments and rating the venue, consumers can choose to record their visits by “check-in” via the LBS mobile app while they are physically present at the venue (Grove 2013). Different from traditional offline engagement (e.g., redeeming paper coupons) or online engagement (e.g., Facebook and Twitter), businesses can interact with mobile consumers by facilitating consumer “check-ins” on LBS given their real-time locations.

In particular, the consumer check-in information generated by LBS adds an essential new dimension to our current knowledge on user engagement. As a unique type of user engagement, check-ins can be interpreted from multiple aspects. First, check-ins have a social aspect. A user could check in to a venue so that her friends may visit the same place after observing her check-ins (Qiu et al. 2018). Second, a unique feature of check-ins is the embedded location information which indicates the current geographical status of a user’s recent behavior. Such location information indicates the user’s taste and preferences. The commonality of check-in locations in a network of users is correlated with friendship formation (Lee et al. 2016). Third, check-ins reflect how mobile users interact with nearby businesses given their current location and can help retailers make decisions about the types of promotions/specials to offer to consumers (e.g., newbie special or friend special[1]).

1.1. The Importance of Consumer Check-ins

Due to a surge in check-ins and consumer reliance on LBS, businesses now have enhanced opportunities to engage with mobile consumers by understanding and cultivating their check-in behavior. LBS users explicitly express their interests in a venue by checking in on LBS apps through their mobile devices (D'Silva et al. 2017). Thus, as a new type of user engagement, check-ins capture a venue’s popularity in the sense that lots of people have visited the venue and are likely to announce their visits to their friends. Understanding the popularity of business venues, revealed by check-ins, is particularly important since a large group of consumers routinely use LBS to seek advice for the venues they may visit based on their current locations.

A venue’s check-ins play an essential role in consumers’ decision-making process because check-ins capture the venue’s popularity and reflect how many consumers have engaged with this venue. In the restaurant industry, an essential task for restaurant managers is to predict the venue’s popularity several days in advance to make arrangements for food and staff (Takenaka 2020). The check-ins provide real-time information for customer engagement through the mobile channel since customers must be

[1] Source: https://mashable.com/2011/05/08/foursquare-special/

physically present at venues to check in on LBS (Wang et al. 2015). In addition, consumer check-ins shorten the feedback loop for businesses' operation decisions, such as food purchase and staffing schedules. Therefore, in this mobile era, how to accurately capture consumer engagement (e.g., check-ins) and its popularity on LBS becomes a vital question for business practitioners, especially from the operations management perspective.

1.2. Research Questions and Approach

Despite the enthusiasm and millions of dollars in investments from businesses, there is limited theoretical understanding and empirical investigation of check-ins on LBS. Recent research on LBS has mainly focused on the social network feature of check-ins and identified factors that boost users' check-ins from users' perspectives (Qiu et al. 2018). Little attention has been paid to understanding the check-ins from the forecasting perspective. Furthermore, the field of operations management has not yet studied the opportunities that check-ins offer to improve businesses' operational decisions. In this paper, we take the initial step to address this question. Instead of using various identification strategies to identify factors that cause more consumers to check in at business venues, we explore how business features help to forecast venue popularity captured by consumer check-ins, which will then provide guidance for improving operation decisions. Specifically, we address the following research questions: *(1) what factors are associated with forecasting venue popularity captured by consumer check-ins? And what are their contributions to venue popularity? (2) Compared to the regression-based model, which advanced model (i.e., machine learning and deep learning models) is the best performing model to forecast venue popularity in the LBS context?*

This paper examines these research questions in the context of the restaurant industry in New York City. Many factors play an influential role in customer engagement forecasting, such as promotions, geo-locations, competition conditions, social media comments (See-To and Ngai 2018), and rating stars (Kim et al. 2016). Most of these factors have been examined independently in the previous literature. This paper aims to predict customer engagement in the form of check-ins on LBS and explore the predictive power of the following business characteristics conjointly: price, quality, promotions, online reviews, and competitors. Our data is collected from Foursquare (a leading company in location-based services) and Yelp (a leading online user review service for local retail businesses).

The complex nature of venue popularity forecasting calls for more advanced techniques to achieve improved forecasting accuracy. The usage of advanced predictive models is becoming more popular and necessary for small businesses such as restaurants (Kilimci et al. 2019). Thus, we implement a variety of machine learning and deep learning models with different characteristics and degrees of sophistication, including linear regression, support vector regression (SVR), random forest (RF), recurrent neural network (RNN), and long short-term memory network (LSTM).

1.3. *Findings and Contributions*

We apply the forecasting models to predict a restaurant's popularity and consumer engagement from two aspects: *Daily Check-in* and *Daily User*. *Daily Check-in* captures the total number of daily check-ins, which can help managers make better day-to-day operational arrangements for food preparation and staff scheduling. Because each user can make multiple check-ins, *Daily User* captures the unique number of customers who visit the venue each day. In each case, we use cross-validations to select the hyper-parameters for machine learning and deep learning models. Comparing the out-of-sample forecast errors across various models helps us identify that compared to the regression model, SVR yields the best forecasting performance among all the advanced models we have examined in this paper. Using the variable selection method, we find that the prediction accuracy of the SVR model reaches the highest when we include all the proposed predictors.

To improve the interpretability of our machine learning and deep learning models, in this study, we apply SHAP as a unifying framework to interpret the contribution from each predictor and compare different models examined in this paper. We find that collectively, all the factors increase the venue's popularity, with mayor check-ins, the number of reviews, and nearby competitors as the top three contributors. Furthermore, using four randomly selected venues to demonstrate the local interpretability, we show that the contributions from each factor vary significantly across different venues.

Our work makes three novel contributions to the operations management literature. First, we are among the first to conceptually and empirically study the role of check-ins in LBS from the business operations perspective. Our findings have theoretical and practical implications for understanding and predicting venue popularity in LBS. Second, using SHAP value, our research quantifies the importance of the information provided by LBS to improve venue popularity forecasting and enhance the model interpretability so that decision-makers can better comprehend the findings from complex predictive models. Finally, we provide empirical evidence on which model performs the best in terms of venue popularity forecasting in a location-based services context. Interestingly, we find that among the models we have examined, support vector regression outperforms all the other models with the highest prediction accuracy, even better than LSTM. This is consistent with the findings from previous literature that deep learning models do not always maintain an edge over other methods for most problems (Rudin and Carlson 2019). Our finding suggests that managers should be cautious when they implement models that are too complicated to understand. Instead, they should aim to construct an interpretable model that can achieve the same level of accuracy as the complex models. Our findings are robust for different customer engagement measures, different training and testing periods, and different prediction horizons.

2. Literature Review

We position our work in the context of the related technology (LBS), consumer engagement, and methodology (prediction). We first discuss the two main types of location related technologies – location tracking and location aware technologies, and discuss how our work adds to this literature. LBS facilitates consumer check-ins, which are a form of consumer engagement. Thus, we situate our work in the context of prior work in consumer engagement. Finally, in terms of methodology, we review the literature on prediction of popularity of online content, and discuss how our work on predicting check-ins adds to this literature.

2.1. Location-Based Services and Its Applications in Operation Management

Location-based services can be categorized into two types: location-tracking services and location-aware services. Location-tracking services provide a user's whereabouts to other users, while location-aware services offer user services (e.g., directions, ads) relevant to the location they are presently at. An example of location-tracking services is RFID, in which location data is used to increase management efficiency and has been examined in previous operation management literature (Bradley et al. 2018). A typical location-aware service is a mobile app that helps the location-specific store advertisements being sent to nearby consumers, such as Foursquare.

Launched in 2009, Foursquare is a leading location-aware service in the LBS market. Users can "check-in" at business venues using their mobile devices by selecting from a list of venues located nearby. The growth of Foursquare registered users and check-ins has been remarkable. Foursquare has more than 60 million registered users, and at least 50 million of them are active on a monthly basis. On average, Foursquare users contribute 9 million daily check-ins[2]. Over 1.7 million businesses are using Foursquare, and the market size of LBS is predicted to be $183.81 billion by 2027 (Gaul 2020).

Besides the tremendous popularity in the business world, LBS has also drawn attention from the academic world. Researchers have begun to explore this emerging field from multiple perspectives, such as the effectiveness of the mobile promotions delivered through LBS (Fang et al. 2015), the performance of geo-fencing (Ho et al. 2020), the relevant privacy concerns when using LBS (Xu 2012), and the impact of social network structure on consumer decision making (Qiu et al. 2018). In addition, computer science researchers have examined location-based services in different contexts: (1) exploring how to use the spatial and temporal information provided by location-based services to make a better location prediction (Gao et al. 2012); (2) understanding users' mobility pattern revealed by their activities on location-based services to provide insights for recommender systems (Cho et al. 2011); (3) studying the venue popularities on location-based services (Li et al. 2013).

[2] Source: https://review42.com/resources/foursquare-statistics/

From the aforementioned discussion, we notice that the business value of check-ins is largely neglected in previous literature. In this paper, we posit that check-ins, as a measure of customer engagement on LBS, can be used to represent venue popularity, and by accurately forecasting it we can provide valuable operational decisions for business venues. Previous LBS literature has mainly focused on identifying causal factors related to consumer check-ins. Little attention has been paid to understanding consumer check-ins from the forecasting perspective. Thus, our paper extends the LBS literature by (1) identifying a group of business features that help to forecast venue popularity captured by consumer check-ins, (2) quantifying their contributions to the forecast accuracy, and (3) identifying the forecast model that fits LBS context the best. The business features provided by LBS include customer reviews, check-ins from the most dedicated customers, promotion, competitors, and promotions offered by competitors.

2.2. Check-ins as Customer Engagement

In this section, we briefly review the literature on customer engagement on LBS and explain why consumer check-ins could be treated as a measure of customer engagement which helps to capture venue popularity. Engagement is defined as "the intensity of an individual's participation in and connection with an organization's offerings and organizational activities, which either the customer or the organization initiates" (Vivek et al. 2012). Playing a critical role in LBS, customer engagement can be viewed as specific interactive experiences between mobile consumers and the business venue (e.g., checking in a venue and commenting on a venue), and other members of the LBS (e.g., becoming friends on LBSN and reading comments left by friends). To examine the impact of check-ins, we study consumer engagement behavior as the outcome for two reasons. First, increased engagement has been linked to increases in venue popularity, purchase frequencies, profitability, and customer satisfaction (Goh et al. 2013, Kumar and Pansari 2016). Second, both theoretical and empirical understanding of factors related to consumer engagement in LBS is still limited and remains a research gap that needs to be filled.

There are two types of engagement behaviors in location-based services: check-ins at a venue and commenting on a venue, both of which are different ways to engage a business on LBS, and both have been used to capture the overall engagement in similar contexts (Lee et al. 2018). In this paper, we focus on consumer check-ins instead of commenting because (1) check-ins are much larger in volume and therefore can have a cumulatively greater impact; (2) the impact of check-ins has been only examined from a social network perspective (Lee et al. 2016), few studies have examined the check-ins from a consumer engagement perspective and used it as a venue popularity measure; (3) commenting has been examined extensively in previous literature (Bai et al. 2020), and specifically, the problem of predicting the popularity of reviews on Foursquare has been examined before (Vasconcelos et al. 2015). However, no concrete empirical evidence on the value of check-in as consumer engagement has been documented in the literature, and we aim to fill this gap in the current study.

2.3. Popularity Prediction of Online Content

Popularity is a measure of content quality for content providers and a way to filter information for content consumers. Predicting the popularity of online content is valuable because of its immediate practical implications: content providers/platforms can better highlight the most popular content, online advertisers can propose more profitable monetization strategies, and online readers can filter a huge amount of information more quickly (Tatar et al. 2014). Thus, a stream of literature has tackled the popularity prediction of online content from various perspectives.

Two types of online content – videos and articles – have been the focus of research on popularity prediction (Bandari et al. 2012, Pinto et al. 2013, Szabó and Huberman 2008, Tatar et al. 2014). Tatar et al. (2014) focused on news articles and proposed models to rank articles based on their predicted popularity captured by the number of comments. Bandari et al. (2012) constructed a multi-dimensional feature space derived from properties of online articles to examine the efficacy of these features served as predictors. Furthermore, scholars have also made improvements in the popularity prediction for video content. Szabó and Huberman (2008) proposed a log-linear model to predict the long-term popularity of online content (captured by the total number of views) from early measurements of users' access to YouTube and Digg. Built on Szabo and Huberman's model (2008), Pinto et al. (2013) proposed a multivariate linear regression model to incorporate information about historical patterns. Finally, in the context of LBS, few papers investigate the popularity prediction of micro-reviews, named as "Tips", provided by Foursquare (Vasconcelos et al. 2015).

Based on the above discussion, among the various types of content provided on Foursquare, the role of consumer check-ins is left under-explored. The popularity of venues on Foursquare is captured by the "check-ins" received for each venue. By accurately predicting which type of businesses attract more visits (i.e., more check-ins), we can provide insights for daily operation decisions (e.g., staff scheduling) and the effectiveness of promotion strategies launched on LBS. In addition, when predicting the popularity of reviews, previous studies have shown that complex machine learning models do not always outperform linear regression models (Vasconcelos et al. 2014). Thus, it is unclear which type of predictive model yields the best prediction accuracy, and we aim to address this issue by comparing linear regression, machine learning models, and deep learning models.

3. Data and Forecasting Methods

3.1. Data Collection

Most of our data is collected from Foursquare, an LBS application that can be used on smartphones. Launched in 2009, it is extensively used worldwide, with an average of about 9 million check-ins per day and supporting many different languages. Foursquare provides consumers an opportunity to explore nearby business venues and offers businesses a channel to reach out to the customers who are in the vicinity. The idea is that people would use their mobile devices to interact with their environment. Subscribers to Foursquare can indicate their interests in a restaurant when they are physically present at or close to the restaurant by 'checking in' to the Foursquare application using their smartphones. Users can check-in at venues using their mobile devices by selecting from a list of venues located nearby.

The data collection process is as follows: (1) we first create a geogrid for NY restaurants and get their geographic information from Foursquare; (2) then we use restaurants' geoinformation to collect their business characteristics from Foursquare and Yelp. In the first step, to make sure our data is representative and to avoid the sample selection bias issue, we do not set any parameters to choose samples. Instead, we obtain data for all the restaurants that are available from Foursquare and Yelp through their API during our data collection period. In the end, we collected publicly available data for 1515 restaurants in New York City from November 28th, 2011, to July 9th, 2012, for 224 days. We only focus on these 1515 restaurants that we found on the first day of our data collection. We do not monitor any new restaurants that joined Foursquare after that. Table 1 reports the summary statistics of our data set. Table 2 presents the correlation matrix, and we find no multi-collinearity issues between our variables.

[Insert Table 1 and Table 2 Here]

Foursquare users can "check-in" to a restaurant on the app using their smartphones when they are physically present at or close enough to the restaurant. As shown in Table 1, *Daily Check-in* is the number of check-ins a restaurant receives each day during the data collection period. It is the primary dependent variable in our application because it helps measure the venue popularity captured by Foursquare. To control the possibility that early adopted restaurants may have more check-ins, we have set the first-day check-in count to be zero for all restaurants and then calculated the daily check-in by focusing on those check-ins generated during our data collection period. In our study, to capture customer arrivals, *Daily Check-in* is treated as a count variable with the zero-inflated negative binomial distribution. Since a customer may check-in more than once at a given restaurant, we also calculate the number of unique users who check-in at a restaurant for each day, and it is presented as *Daily User* in Table 1. We use *Daily User* as another measure of venue popularity in the robustness check and demonstrate that the prediction performance remains consistent across the two accuracy measures. We have conducted multiple checks to validate the distribution that best fits *Daily Check-in* and *Daily User* (see E-Companion B for details).

Most restaurants have dedicated customers who visit the venue more frequently than other customers. The customer who visits a restaurant the most is labeled as the "Mayor" of that restaurant on

Foursquare. We define *Mayor Check-in* as the number of check-ins a mayor has made to capture the effect of this dedicated customer on the venue popularity prediction. Since the mayorship depends on the number of check-ins a consumer has made, today's mayor may lose the title tomorrow if another consumer has made the highest number of check-ins at the same venue. Thus, the mayor is a title that refers to a unique group of consumers, those who have made the highest number of check-ins at the venue. *Mayor Check-in* captures the contribution from this special group of consumers instead of a single/specific consumer.

Traditionally, consumer loyalty is related to consumer lifetime value (Zhang et al. 2010), repeated purchase rate (East and Hammond 1996), and engagement level (So et al. 2016). Inspired by previous literature, we treat *Mayor Check-in* as a measure of consumer loyalty for the following reasons. First, as a unique type of frequent visitor, mayors can bring a higher customer lifetime value than other consumers who have visited the same venue. Second, by definition, mayors are the group of consumers who have the highest repeated visit rate. They are consumers who come back repeatedly to dine in at the restaurants in our data set. Third, mayors are the group of consumers who are more engaged than other consumers. Furthermore, as pointed out by Rothenberger et al. (2008), there are two widely used customer loyalty indicators: the likelihood of recommending the firm to others and the likelihood of reusing the service. This provides further evidence that we can treat *Mayor Check-in* as a measure of consumer loyalty since it incorporates both of these aspects.

Restaurants who subscribe to Foursquare can use it as a free channel to offer promotions and discounts to their customers, and these promotions/discounts are labeled as "Specials." *Specials* is used to indicate the number of promotions a restaurant offers on this mobile app. Specials can be tailored based on different characteristics, such as the type of customers the business wants to attract, the discount rate the business offers, and the format of rewards. With that being said, quantifying the special offer is almost an impossible job given its variety. Therefore, we focus on the number of specials a restaurant offers to its consumers instead of the magnitude or types of such discounts. Like many other social media platforms, customers can provide review comments about a restaurant on Foursquare after their visits. Defined as the number of reviews a restaurant receives each day, *Daily Review* is a measure to capture the word-of-mouth (WOM) effect from the review quantity perspective. We have also examined the WOM effect from the review valence perspective, the details of which are reported in E-Companion K.

Two variables are utilized to describe the competitive environment faced by each restaurant in the context of location-based services. First, *Competitor Total Number (CTN)* is the number of competitors a restaurant has on Foursquare, which is directly collected from Foursquare. For a focal restaurant, its competitors are identified by two criteria on Foursquare: (1) both competitors and the focal restaurant must belong to the same type of cuisine; (2) competitors are located within a 3-mile radius of the focal restaurant. There are maximally five competitors identified by Foursquare per restaurant. Second, *Competitor Special Total Number (CSTN)* is defined as the number of specials offered by a restaurant's competitors. This is not directly provided by Foursquare, but we calculate this variable using competitors' unique identification numbers.

Yelp_PriceRange and *Yelp_Rating* are obtained from Yelp and refer to the expensiveness and quality of a restaurant. *Yelp_PriceRange* has values from one to four, with one indicating that the average expense per customer is below $10, two indicating that it is between $10 and $30, three indicating that it is between $30 and $60, and four indicating that it is above $60. *Yelp_Rating* is a regular star-rating scheme with one star indicating the lowest rating and five stars indicating the highest rating a restaurant can get from customers.

3.2. Forecasting Models

Panel or longitudinal data, in which repeated observations are available for each sampled object, offers a rich opportunity for prediction. We observe the different paths over time that a response variable may take across subjects. Such data are often seen in many applications, such as business and bioinformatics (Sela and Simonoff 2012). An analyst might be interested in two types of tasks given a panel data set: modeling and prediction, in which prediction is a more immediate need given a panel data structure. We focus on the panel data structure we collected (1515 restaurants with 224 days of observations on venue popularity revealed by Foursquare) and examine the prediction power of various prediction models with different levels of complexity.

In the machine learning literature, there are at least four main families of "supervised" learning algorithms: (1) logical models (i.e., decision trees); (2) linear combinations of trees, stumps, and other kinds of features (i.e., random forest, additive models); (3) case-based reasoning (i.e., SVM with different kernels); (4) iterative summarization (i.e., neural networks) (Rudin and Carlson 2019). Logical models, such as decision trees, do not fit in our research context given the nature of our target variables (*Daily Check-in* and *Daily User*). Thus, we implement six widely adopted statistical algorithms chosen from the other three families of supervised learning algorithms: panel models and random forest from linear combinations, support vector regression from case based-reasoning, and recurrent neural network and long short-term memory networks from iterative summarization.

The panel models we adopted are the Pooled OLS model and the generalized linear mixed-effect model with zero-inflated negative binomial distribution assumption. Support vector regression (SVR) has been used in various research contexts, such as vehicle crash prediction in highway safety (Li et al. 2008), demand forecasting in the supply chain (Kilimci et al. 2019), and clinical trials in healthcare (Du et al. 2015). Du et al. (2015) propose a novel longitudinal SVR algorithm that takes advantage of this popular machine learning method and models the temporal nature of longitudinal data by considering observational dependence within subjects. Li et al. (2008) demonstrate the superiority of SVR (with radial basis function as kernel function) over Negative Binomial models on predicting vehicle crashes. This paper adopts SVR and implements it on our venue popularity (i.e., customer check-ins) prediction problem. As for the random forest, we adopt the random forest model proposed by Hajjem et al. (2014) and modify it to fit our panel data analysis and capture the potential nonlinear effects in our venue popularity prediction problem.

Deep neural networks are state-of-the-art methods that provide a framework to model complex, nonlinear interactions in large datasets. Therefore, they are naturally suited to hierarchical data analysis, such as panel data (Falissard et al. 2018). Recurrent neural networks (RNNs) are a family of deep neural networks specializing in analyzing panel data (Goodfellow et al. 2016). Long short-term memory network (LSTM) is an artificial RNN architecture used in the field of deep learning. It is known for its superiority in capturing the long-term dependencies compared to traditional RNNs (Hochreiter and Schmidhuber 1997). In this paper, we test the prediction performance of RNN and LSTM. E-Companion A1 describes the details for each forecasting model evaluated in this paper.

Furthermore, we would like to highlight that all the relationships examined in this study are associations instead of causality. The focus of this paper is to forecast venue popularity. Thus, our goal is to identify factors that help forecast future venue popularity captured by consumer check-ins. Following the previous literature (Cui et al. 2018), we have proposed a standard forecasting framework in the next section: (1) training/testing dataset and cross-validation, (2) out-of-sample evaluation, (3) L-Day-Ahead forecasting, and (4) model interpretability.

3.3. Forecasting Framework

Given that customers may visit a restaurant more than once, each restaurant's customer arrival can be captured by two variables: the number of "check-ins" and the number of unique customers. *Daily Check-in* reflects the total number of visits a restaurant has each day, whereas *Daily User* captures the unique number of customers who visit a restaurant each day. Given the differences between the two measures, in this paper, we use both of them to capture each restaurant's popularity on Foursquare and examine how we can improve the venue popularity prediction using models with varying complexity. We report and discuss the prediction results using *Daily Check-in* as the dependent variable in the following section. *Daily User* is used as a robustness check and the prediction results are summarized in E-Companion E.

Training and Cross-Validation

To evaluate the prediction model performance, we divide the data into two parts: 80% of the original data set is used as the in-sample training set, and the remaining 20% is used as the out-of-sample testing set. The training set is used to train the model and the testing set is used to evaluate it. We denote the training period as [1, 2, …, T], and the out-of-sample testing period as [T+1, T+2, …, T+N], where N is the number of days in the out-of-sample testing period. In our study, the in-sample training set has 182 days (from 2011-11-28 to 2012-05-28), T=182, and the out-of-sample testing set has 42 days (from 2012-05-29 to 2012-07-09), N=42. To validate the robustness of our findings across different time frames, we change the size of the training set (70% of the original data) and testing set (30% of the original data) so that the starting time of the training set and testing set varies accordingly. The prediction results of this robustness check are summarized in E-Companion F.

Cross-validation is used to choose the value of the hyperparameter that gives the best performance. There are two types of cross-validation: traditional K-fold and stacked K-fold. For traditional K-fold, the training set is randomly partitioned into K equal-sized subsets. Of the K subsets, a single subset is retained as the validation data for testing the model's performance and the remaining K-1 subsets are used as training data. Instead of randomly shuffling all data points and losing their order, stacked K-fold split the data in order. We demonstrate the difference between the traditional K-fold and stacked K-fold in Figure 1.

In contrast to the traditional K-fold, which uses all the data at each iteration, the stacked K-fold only uses the data from past iterations for training and new data for testing. The process of splitting the dataset into training and testing portions repeats for K times. The size of the testing data remains fixed but the training subset changes through the original dataset, and the remainder is used as a training dataset in every fold shown in Figure 1. When all iterations are over, we choose the model with the highest performance, which may not even be the last one. In this paper, we use the training data set for cross-validation and apply the stacked K-fold method where K=4.

[Insert Figure 1 Here]

Out-of-Sample Evaluation

When we construct predictions for the out-of-sample testing period, we use the trained model and make predictions using the most recent information up to the time of the forecast. Specifically, to predict venue popularity for day T+1, we use the model that is estimated using the training data from day 1 to day T along with the values of input variables for day T+1. Thus, for the linear algorithms and machine learning classification algorithms, we use the values of predictors in the testing sample to make the prediction. In contrast, such data is not needed for deep learning algorithms. For RNN and LSTM, after using the training set to train our model, we use the most recent 30 days up to Day T to predict Day T+1 without knowing the predictors' value of Day T+1.

To select the best model, we choose two metrics that are less affected by the skewness. The first one is the mean absolute error (MAE), which is calculated as the mean of the absolute errors. It is written as $AE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$. This method is chosen as it gives general information on the size of the errors.

The second one is the root mean squared error (RMSE), which is calculated as the square root of the mean of the squared error. It is written as $RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$. This method is chosen as it gives information on the size of the error while penalizing larger errors.

L-Day-Ahead Forecasts

So far, we have discussed how to construct venue popularity prediction for the next following day. In practice, businesses need lead time to adjust their production and operational decisions. Thus, they often need to forecast more than one day. Consequently, we construct an L-day-ahead forecast. We estimate how previous business information up to day t impacts the venue popularity on day t+L. We also verify whether the prediction accuracy is consistent over longer lead times, ranging from one to seven days. Following the suggestions by Cui et al. (2018), we examine the one-day, three-day, five-day, and seven-day-ahead forecast (L=1, 3, 5, and 7).

Model interpretability – SHAP value

Understanding why a model makes a certain prediction is as important as the prediction accuracy in many situations, and the growing tension between accuracy and interpretability has motivated the development of methods that help users interpret predictions (Lundberg and Lee 2017). Higher interpretability of the model helps managers understand predictive models better and apply them more confidently. It also helps managers communicate the analytical rationale for their decisions to stakeholders more convincingly. Interpretable machine learning has thus become a research area that has been attracting a lot of attention. Multiple methods have been proposed in the literature to enhance the interpretability of predictive models (Biecek 2018). Generally speaking, there are two types of interpretable approaches, (1) local interpretability, which focuses on personalized interpretation (i.e., local interpretable model-agnostic explanations (LIME)), and (2) global interpretability, which summarizes prediction models on a population level (i.e., SHAP) (Stiglic et al. 2020). Given that our focus is on the restaurant industry in Manhattan rather than a single venue, we decide to apply SHAP to quantify the contributions of our predictor variables, and to interpret and compare the predictive models examined in this paper.

Proposed by Lundberg and Lee in 2017, SHAP uses Shapley values from game theory to explain a specific prediction by assigning an *importance value* (SHAP values) to each feature that has the following properties: (1) local accuracy; (2) missingness; (3) consistency (Antwarg et al. 2021, Lundberg and Lee 2017). Lundberg and Lee (2017) demonstrate that SHAP is better aligned with human intuition than previous methods because it has these properties.[3] A detailed explanation of SHAP is provided in E-Companion A2. We use the SHAP Python module to calculate the feature importance values (SHAP

[3] It is worth noting that SHAP value does not provide any information about causality.

value) and create two plots: (1) a global feature importance plot that shows the collective feature importance for each predictor and the correlation between predictors and target variable; (2) a local feature importance plot that shows how the contributions of our predictors vary across four randomly selected individual venues.

4. Prediction Results and Model Interpretation

In this section, to predict venue popularity revealed by Foursquare, we focus on *Daily Check-in* and compare the prediction performance of the proposed algorithms and discuss the findings. To identify the best value for the hyperparameters used in our models, we perform the 4-fold stacked cross-validation and choose the following configuration for each model.

SVR: Use linear kernel as the kernel function.[4]

MERF: The number of trees: 100, node size: 5.

RNN: The number of hidden states: 16, batch size: 2000, epochs: 10.

LSTM: The number of hidden states: 16, batch size: 2000, epochs: 10.

4.1. Venue Popularity Prediction Results

We implement six algorithms discussed above (in Section 3.2) to predict venue popularity revealed by Foursquare. To compare their prediction performance, we adopt the bias-variance trade-off framework in machine learning (Friedman et al. 2001). Prediction error contains three parts: irreducible error, bias, and variance. Thus, to minimize the prediction error, we need to minimize bias and variance. In general, less flexible models have higher biases, whereas more flexible models have higher variances. A model with a lower variance usually comes with a higher bias and vice versa. In previous literature, the in-sample errors are used as proxies for biases, and the out-of-sample errors are used as proxies for variances (Friedman et al. 2001).

Models can be classified into three groups based on their errors: (1) under-fitting, when both in-sample and out-of-sample errors are high, which means the model has a high bias; (2) over-fitting, when the in-sample error is low and out-of-sample error is high, which means the model has a high variance; (3) good fitting, when the out-of-sample error is only slightly higher than the in-sample error and both of them are low. Table 3 summarizes the in-sample and out-of-sample RMSE and MAE for each of the algorithms. We find that SVR outperforms both the linear regression algorithms and the deep learning algorithms based on the in-sample and out-of-sample error comparison. We do not spot any serious over-fitting or under-fitting issues with our machine learning models.

[Insert Table 3 here]

[4] To alleviate any concerns about nonlinear combinations of our features, we have compared the radial basis function kernel with the linear kernel. We find that the linear kernel gives us the best prediction performance. We further normalize the predictors to address any concerns that SVR is very sensitive to the range of features. We find that our results are consistent after trying different kernel functions and normalization. Details can be found in E-Companion G.

Figure 2 presents the out-of-sample RMSE and MAE performance for the six algorithms examined in this paper. For each model, the darker column represents RMSE, and the lighter column represents MAE. After comparing the RMSE and MAE for the 1-day-ahead forecast, we find that among the six algorithms, SVR is the best in terms of forecast accuracy. SVR outperforms all other algorithms in predicting venue popularity revealed by Foursquare[5]. This finding benefits Foursquare and other location-based services in various ways. Specifically, at the platform level, Foursquare can better identify the most popular/trending venues in various regions and provide better restaurant-recommendations to Foursquare users. At the venue (restaurant) level, this accurate prediction on venue popularity among consumers may help to improve its daily operation decisions (i.e., procurement and staff scheduling), which then enhances the venue's business performance and attracts more consumer engagement with both the venue and the platform (i.e, Foursquare).

[Insert Figure 2 here]

It is interesting to observe that SVR outperforms linear regression. One possible explanation is that SVR uses a different loss function for the process of minimizing error. Linear regression uses the squared error loss for each training example, also known as the L2 Loss. However, for linear-support vector regression, assuming linear parameterization $f(X, \omega) = W \cdot X + b$, the ε-insensitive loss function is defined as: $L_\varepsilon\big(y, f(X, \omega)\big) = \max\,(|y - f(X, \omega)| - \varepsilon, 0)$. In SVR, observations within the threshold of ε will be ignored, and only the observations outside of the ε range contribute to the final cost. The advantage of SVR over linear regression is that linear regression models do not approximate the underlying data generation process very well for high-dimensional problems due to over-fitting issues (Cui et al. 2018).

We have also found that SVR is superior to all our deep learning models (RNN and LSTM) for this venue popularity prediction. Neural networks, such as RNN and LSTM, typically have complicated architectures and optimization procedures, and they are very difficult to tune. Furthermore, neural network performance depends on the initialization of the randomly-generated parameters. Thus, the poor performance of these neural network-based models may be because our prediction problem lacks the complex nonlinear relationships which are better captured by neural networks. Furthermore, although deep learning models (i.e., RNN and LSTM) are superior to machine learning models like SVR in some contexts (Zhang et al. 2018), there are a vast set of problems for which neural networks do not have an advantage over other methods (Rudin and Carlson 2019). In particular, when the predictors have an inherent meaning (i.e., age, gender), most machine learning methods perform as well as deep learning models if tuned properly. As suggested by Rudin and Carlson (2019), for regression data with inherently

[5] The estimation time on the testing set is 13.5 minutes for SVR, 1.6 hours for MERF, and around 12 minutes for RNN, and 17 minutes for LSTM.

meaningful covariates, we should try different algorithms and choose the simplest or most meaningful one among the ones that perform similarly after parameter tuning. Thus, after cross-comparing the performance of the six algorithms chosen from three families of supervised learning algorithms, we conclude that SVR is the best predictive algorithm in our research context.

As discussed previously, we implement our analysis to forecast venue popularity using different L-day-ahead forecasts. Table 4 summarizes the daily-level out-of-sample forecast accuracy for all algorithms we examined. For both the RMSE and MAE, the forecast error increases slightly as the lead time increases. This is what we would expect since the further ahead we forecast venue popularity, the less likely future popularity depends on historical observations due to future uncertainty. Deep learning methods such as RNN and LSTM have a relatively sizeable increase in prediction errors because their predictions do not use predictors' future values, which leads to higher future uncertainty.

[Insert Table 4 here]

Ideally, we would like to compare our L-day-ahead forecast to the restaurant managers' judgment-based forecast to demonstrate the superiority of the advanced forecasting algorithms. Due to the data collection limitation, we do not have data for such a judgment-based prediction. Therefore, we use the RMSE and MAE of the generalized mixed-effect model (GLME) only as a benchmark and compare the machine learning models and deep learning models with this regression-based method. Comparing the difference between the GLME model and our machine learning and deep learning models, we capture an estimate of the value of applying advanced forecasting methodology. When the lead time is one day, the forecast with the SVR model generates an RMSE of 3.609. In contrast, the baseline model (the fixed-effect panel model) gives a higher RMSE of 5.295, with a relative forecast accuracy improvement of 31.8% calculated as: $\frac{Baseline\ RMSE - SVR\ RMSE}{Baseline\ RMSE}$.

The benefits of using a more advanced forecasting methodology, captured by the relative RMSE improvement, are 3.8% from RNN, 6.8% from LSTM, 11.5% from MERF, and 31.8% from SVR. When we increase the lead time, the improvement captured by SVR and MERF remains the same, whereas the improvement captured by RNN and LSTM decreases and even drops below zero when we predict the venue popularity for more than five days ahead. The relative improvements of using advanced forecasting algorithms are summarized in Figure 3.

[Insert Figure 3 here]

To examine the forecasting power of the temporal factors, we have added weekday dummies into our forecasting models, and the results show that weekday dummies help to reduce the forecasting errors. However, such a decrease is small and negligible compared to the case without the weekday dummy as shown in E-Companion I. Furthermore, we have also included the interaction term, *Specials***CTN*, as a predictor in our forecasting models. We have found similar results that the interaction term helps to reduce the forecasting errors, though these effects are very small and almost negligible, as shown in E-Companion J.

4.2. The Importance of Location-Based Service Features

In this research context, location-based services provide a channel for businesses to attract and engage the customers in their vicinity in a more proactive way. Businesses could send out customized promotions and rewards to targeted customers and inform customers who are new to the vicinity about their service and products. These customers can check in at this venue through the app, and engage with both the business and the app. The check-ins made by customers not only demonstrate customer engagement but also reveal the venue's popularity on the platform.

As for customers, by looking through the business's profile page, they can check not only the service or products, its popularity, and promotions but also other customers' reviews and ratings about this venue along with other potential choices within a 3-miles radius. All these factors contribute to customers' engagement decision. Therefore, venue popularity is affected by a set of factors, including price, rating, potential competitors nearby, the impact of its loyal customers, etc. We next study how vital these factors are and the effect of each factor on forecasting accuracy. We do so by (1) using the variable selection method to identify the "best" subset of predictors; and (2) applying SHAP value to interpret results we get from the machine learning and deep learning models.[6] Specifically, we present SHAP value for global interpretation and also demonstrate its local interpretation ability using three individual cases. We focus on the SVR model in the following discussion and summarize the SHAP value for other machine learning and deep learning models in E-Companion H.

Variable Selection Analysis

We apply the variable selection method in this paper to capture each predictor's contribution to the prediction accuracy. Variable selection is widely used in predictive analysis to select a subset of variables that allow the construction of the best predictor (Reunanen 2003). The benefits of using variable selections include: (1) improving prediction accuracy through the exclusion of irrelevant variables; and (2) a better understanding of the prediction problem by knowing which variables are relevant. Stepwise procedures, both forward and backward, are used to add or remove variables from a model sequentially. We use backward elimination, the simplest and most accessible of all variable selection procedures, to select the best set of predictors. We start with the full model, including all the predictors we are interested in, and keep eliminating variables one by one until the prediction accuracy is no longer improved.

[Insert Table 5 Here]

[6] The linear regression estimation results are reported in E-Companion C.

In this section, we focus on the SVR model due to its best prediction performance.[7] The variable selection results for the SVR model are presented in Table 5. We observe that the prediction accuracy of the SVR model reaches the highest when we include all the predictors in the analysis, and this finding remains consistent for all other prediction lead times. Furthermore, we also notice that the RMSE and MAE have very small variations (almost negligible) when we apply the backward elimination on the SVR model. Thus, we decide to keep all the variables in our prediction model when we explore the explainability of the SVR model in the next section.

Model Interpretation – SHAP Values

In this section, we start with the global interpretability of SHAP values and then demonstrate its local interpretability using four individual venues. Traditional variable importance methods only focus on the importance of predictors at the population level. However, SHAP values take one step forward by providing local interpretability for the individual cases, which enables us to pinpoint and contrast the impacts of the predictors for each case.

The collective SHAP values can be used to show each predictor's contribution to the target variable, either positive or negative. We present the contributions from each predictor in Figure 4. All variables are ranked in descending order. Variables on top contribute more to the predictive model than the bottom ones and therefore have higher predictive power. The horizontal bar shows whether the effect of that predictor is associated with a higher or lower prediction, while the color shows whether the predictor is positively (in red) or negatively (in blue) associated with the target variable. From Figure 4, we observe that *Mayor Check-in* has the highest contribution to the venue popularity prediction. A high level of mayor check-in has a high and positive impact on the venue popularity on LBS. Thus, the number of check-ins made by the venue's mayor is a big predictor of the venue's popularity on LBS.

The next important factor is the *Competitor Total Number (CTN)*, which has a positive impact on the venue's daily check-ins. This is a sign of the agglomeration effect. There are two types of spatial dependence, agglomeration, and competition. Agglomeration means that businesses will benefit from locating close to each other (Pancras et al. 2012). Competition means that businesses will get hurt if they are close (Davis 2006). Pancras et al. (2012) studied the agglomeration effect using restaurant data from the same fast-food chain. Inspired by their work, our *CTN* variable is defined similarly, as the number of nearby restaurants that belong to the same cuisine. Given this positive relationship between *CTN* and *Daily Check-ins*, we label it as the agglomeration effect instead of the competition effect. In this interpretation, we are guided by the theoretical approach of Fotheringham (1983), who points out the

[7] The variable selection results for the other algorithms are reported in E-Companion D.

proximity to other destinations as indicators of both agglomeration and spatial competition, where the empirical context determines which of these countervailing forces dominate.

Daily Review captures the impact of the WOM effect from the review quantity perspective. We have found that a higher number of reviews left by consumers has a high and positive impact on the daily check-ins a venue has. Furthermore, *Specials*, *Yelp_PriceRange*, and *Yelp_Rating* have similar positive contributions to the prediction of venue popularity captured by check-ins. Compared to all other predictors, *Competitor Special Total Number* contributes the least to the venue popularity given its lowest SHAP value.

[Insert Figure 4 Here]

Besides global interpretability demonstrated by Figure 4, each venue also gets its own set of SHAP values which significantly improve the transparency of the "black box" types of predictive models (i.e., RNN, LSTM). Each set of SHAP values explains why a venue receives its prediction and the contribution of the predictors. We randomly draw four venues (V1-V4) and demonstrate how each predictor contributes to its venue popularity prediction. It is intriguing to observe that although all the predictors have positive associations with the daily check-ins prediction, each predictor's contribution varies significantly across individual venues.

[Insert Figure 5 Here]

The base value is the mean prediction. In other words, it is the daily check-ins that would be predicted if we do not know any features of that venue. Our analysis shows that, on average, each venue gets around two check-ins per day (the base value). The prediction for the first venue's daily check-in is 1.71. The only predictor that has a positive impact (shown in red) on the venue's daily check-ins is the total number of competitors nearby. In contrast, mayor check-in, price, and rating have negative impacts (shown in blue) on daily check-ins. Quantitively*, Mayor Check-in,* and *CTN* have much higher contributions to the prediction than *Yelp_PriceRange* and *Yelp_Rating*. Specifically, this venue (V1) has five competitors nearby, which is more than the average value of 4. Thus, *CTN* pushes the prediction to the right. The restaurant has three mayor check-ins, a 3.5 rating, and a $15-$30 price range (2), which are all lower than the average means[8], and drive the prediction to the left. For another venue (V3), all the predictors are higher than the corresponding average values, and thus push the prediction on daily check-ins higher to the right and reach almost four check-ins every day. We also notice that the check-ins made by the mayor contribute the most to the prediction for venue 3 (V3), while price, rating, and competitor numbers drive the prediction in a similar marginal way. However, the same three predictors make much stronger positive contributions for venue 4 (V4). Our findings show how the same set of predictors drive

[8] The averages for mayor check-in, rating, and price range are 11, 3.57 and 2.14, respectively.

the target variable into different directions with different forces for individual venues, highlight the importance of local interpretability, and introduce SHAP value to venue managers so that they can better comprehend the results yielded by complex prediction models.

5. Conclusion

With the increasing number of location-based services and customer check-ins at business venues, location-based services have the potential to play a prominent role in attracting more consumer engagement for businesses and increasing their popularity. In this research, we studied the businesses' venue popularity prediction using the customer engagement data (check-ins) on LBS, and have shown how multiple machine learning and deep learning models contribute to this prediction with various improvements. These results apply to the out-of-sample forecast test for both *Daily Check-in* and *Daily User* to capture venue popularity. The results are also robust to different subsamples of data with varying periods of training.

Improving venue popularity prediction can lead to substantial operational benefits in various situations. A good venue popularity prediction can provide accurate information about customer engagement with the venue, which complements demand forecasting efforts and helps managers reduce inventory costs and schedule the staff more efficiently. An accurate venue popularity forecast is especially important for businesses dealing with perishable products and high employee turnover rates, such as the restaurant industry.

In the context of small businesses, the accurate prediction of venue popularity, particularly captured by customer check-ins, requires careful analysis of multiple factors, such as promotional activity, price, online rating, online review, nearby competitors' activity, etc. To quantify the contributions from each factor and enhance the model interpretability, we applied SHAP as a unifying framework to interpret and compare different types of predictive models examined in this paper. The global interpretability of SHAP values suggests that all the factors contribute to the venue popularity prediction with different levels of feature importance. The top three contributing factors (in descending order) are *Mayor Check-in*, *Competitor Total Number*, and *Daily Review*. Specifically, the check-ins made by the venue's mayor are the biggest predictor of the venue's popularity on LBS. A high level of mayor check-ins has a high and positive impact on the venue's popularity on LBS, demonstrating the importance of customer loyalty. Furthermore, the positive impact of *Competitor Total Number* suggests the agglomeration effect: the more competitors nearby, the more customer check-ins a venue will attract, leading to higher venue popularity. The number of reviews posted on Foursquare also drives the venue's popularity and attracts more consumer check-ins, capturing the positive word-of-mouth effect. Then, to demonstrate local interpretability, we have calculated the SHAP values for four randomly selected venues and found that each predictor’s feature importance varies significantly across individual venues, although collectively all the predictors have positive associations with the daily check-ins prediction.

The importance of using cutting-edge machine learning and deep learning models to enhance prediction accuracy has been recognized in previous literature (Choi et al. 2018). The performance of each model varies substantially based on the research context. For instance, the random forest is the best model to demonstrate the predictive power of social media information (Cui et al. 2018), whereas neural networks outperform other sales forecasting models in fashion retailing (Sun et al. 2008). As suggested by Rudin and Carlson (2019), we have compared various machine learning and deep learning models and enriched the venue popularity prediction research by identifying that SVR outperforms other advanced models in the context of location-based services. SVR is a valuable and flexible technique with the ability to deal with the limitations pertaining to distributional properties of underlying variables and the common problem of overfitting. We have observed that SVR is superior to linear regression as it optimizes the parameters for best prediction using cross-validation, which is missing in linear regression. We have also found that SVR outperforms deep learning models that are based on neural networks. This result indicates that the complex non-linearities that deep learning models typically capture may not be a significant feature in our context. Furthermore, previous literature suggested that when the predictors have an inherent meaning (i.e., age, gender), most machine learning methods perform similarly as deep learning models if tuned properly. This may explain the superior performance of SVR as compared to deep learning models in our context, given that our predictors are business characteristics all having inherent meanings rather than raw measurement values (i.e., raw pixel values from images).

Prior work has found that most people using machine learning or deep learning algorithms do not even attempt to create an interpretable model due to the fear that they may need to sacrifice accuracy to gain interpretability (Rudin and Carlson 2019). On the contrary, our analysis shows that it is possible to produce an interpretable model without sacrificing accuracy. Higher interpretability of predictive models increases the likelihood that managers would utilize them for decision-making. This has led to interpretable machine learning becoming an area of focus for recent research (Lundberg and Lee 2017). Among various methods proposed in previous literature, we adopt SHAP to enhance the interpretability of our predictive models. SHAP uses Shapley values from game theory to explain specific prediction results by assigning a feature importance value (SHAP value) to each predictor in the model. To the best of our knowledge, we are among the first in the Operations Management field to implement SHAP value to enhance the interpretability of our prediction algorithms.

Our paper provides several actionable implications for decision-makers such as restaurant managers. First, based on our SHAP analysis, we suggest that restaurant managers should start to attract more customer engagement on LBS by (1) motivating more engagement from their mayors, (2) staying in an area with similar businesses which do not provide many promotions, and (3) encouraging consumers to provide reviews for their venues. These factors reflect the essential contributions of loyal customers, the agglomeration effect, and the word-of-mouth effect in boosting the customer engagement and venue popularity on LBS. Second, managers should be cautious when they implement the models that are too complicated to understand, such as deep learning models. For high-stakes decisions, complex models should not be used unless absolutely necessary. Instead, managers should aim to construct an interpretable model that produces accurate predictions compared to complicated "black-box" models.

Our research could be extended in multiple ways. First, according to review trackers,[9] 45.18% of consumers rely on Yelp.com when deciding on visiting a retail business location. Since our methodologies are tested on data from two separate LBS services – Yelp as well as Foursquare, this provides some evidence for generalizability of our results in the restaurant industry as well as with industries with similar characteristics in the retail service sector. Further research is needed to build on our work to develop accurate models for predicting venue popularity in other sectors. Second, the current study focuses on the restaurant industry and examines mainly short-term forecasts. Our models could be easily extended to other types of businesses, such as shopping malls and theaters, to understand their customer engagement patterns and make more accurate venue popularity predictions to improve business performance.

Third, our paper is only the first step in using the information on LBS to forecast customer engagement and venue popularity. With more data collected from other LBS apps, we could use our model to forecast consumer engagement and venue popularity for productions and services on other LBS apps. We posit that based on future research, our findings could be easily generalized to other LBS settings for the following two reasons: (1) the predictors we used in this research are common business characteristics that can be easily identified on multiple LBS apps (i.e., Yelp). We suspect that consumer loyalty, the agglomeration effect, and the WOM effect would still be the top contributors for venue popularity with a different order. (2) Since all the predictors in our models have inherent business meanings (see discussions in Section 4.1), we posit that SVR still has a significant chance of outperforming other machine learning and deep learning models in other LBS contexts.

Limitations

While taking an initial step to forecast venue popularity in the LBS context, this study is constrained by some data limitations. First, we acknowledge that temporal and spatial information (i.e, day of the week and spatial proximity to nightlife establishments) are essential factors in forecasting venue popularity. Ideally, we should incorporate both factors, but we have only included the weekday

[9] https://www.reviewtrackers.com/reports/online-reviews-survey/

dummy as our temporal factor due to missing data on the spatial factor. This issue is one limitation and a priority for future work. Furthermore, due to the data collection challenges, we are not able to incorporate other types of spatial information (i.e., the total number of restaurants within a 5-mile radius) in the current paper, but it would be a good robustness check to include in the future. Second, we currently do not separate restaurants into different types such as chain and independent restaurants. A natural way to extend this research is to examine the role of the restaurant types (chain vs. independent) on forecasting venue popularity. Third, although Foursquare has a broad user base and has become popular for many years, it is relatively more appealing to the young and urban demographic group, which differs from the general U.S. population. This limitation puts constraints on Foursquare's predictive power for customer engagement and venue popularity that is targeted for other demographic groups.

Despite the abovementioned limitations and new directions, this study, to the best of our knowledge, is among the first to forecast venue popularity in the LBS context using various machine learning and deep learning algorithms. Overall, this study generates executable guidance on how businesses can effectively leverage LBS to make better operational decisions such as procurement and staff scheduling based on venue popularity prediction.

Acknowledgment

The authors thank Professor Subodha Kumar, the senior editor, and anonymous reviewers for their valuable and constructive suggestions which have significantly improved this research. Ram Gopal gratefully acknowledges support by the Gillmore Centre for Financial Technology at the Warwick Business School.

References

Antwarg, L., R. M. Miller, B. Shapira, L. Rokach. 2021. Explaining anomalies detected by autoencoders using Shapley Additive Explanations. *Expert Systems with Applications* **186** 115736.

Bai, X., J. R. Marsden, W. T. Ross, G. Wang. 2020. A Note on the Impact of Daily Deals on Local Retailers' Online Reputation: Mediation Effects of the Consumer Experience. *Information Systems Research* **31**(4) 1132-1143.

Bandari, R., S. Asur, B. Huberman. 2012. The Pulse of News in Social Media: Forecasting Popularity. Presenation, the 6th International AAAI Conference on Weblogs and Social Media, Dublin, Ireland.

Biecek, P. 2018. DALEX: explainers for complex predictive models in R. *The Journal of Machine Learning Research* **19**(1) 3245-3249.

Bradley, R. V., T. L. Esper, J. In, K. B. Lee, B. C. Bichescu, T. A. Byrd. 2018. The Joint Use of RFID and EDI: Implications for Hospital Performance. *Production and Operations Management* **27**(11) 2071-2090.

Cho, E., S. A. Myers, J. Leskovec. 2011. *Friendship and mobility: user movement in location-based social networks*. Presentation, the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, August 21-24, San Diego, California, USA.

Choi, T.-M., S. W. Wallace, Y. Wang. 2018. Big Data Analytics in Operations Management. *Production and Operations Management* **27**(10) 1868-1883.

Cui, R., S. Gallino, A. Moreno, D. J. Zhang. 2018. The Operational Value of Social Media Information. *Production and Operations Management* **27**(10) 1749-1769.

D'Silva, K., A. Noulas, M. Musolesi, C. Mascolo, M. Sklar. 2017, 2017. *If I build it, will they come?* Presentation, the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Novembe 7-10, Redondo Beach, CA USA.

Davis, P. 2006. Spatial competition in retail markets: movie theaters. *Rand Journal of Economics* **37**(4) 964-982.

Du, W., H. Cheung, I. Goldberg, M. Thambisetty, K. Becker, C. A. Johnson. 2015. A Longitudinal Support Vector Regression for Prediction of ALS Score. *IEEE International Conference on Bioinformatics and Biomedicine* **2015** 1586-1590.

East, R., K. Hammond. 1996. The erosion of repeat-purchase loyalty. *Marketing Letters* **7**(2) 163-171.

Falissard, L., G. Fagherazzi, N. Howard, B. Falissard. 2018. Deep clustering of longitudinal data. *ArXiv*, *abs/1802.03212*.

Fang, Z., B. Gu, X. Luo, Y. Xu. 2015. Contemporaneous and Delayed Sales Impact of Location-Based Mobile Promotions. *Information Systems Research* **26**(3) 552-564.

Fotheringham, A. S. 1983. Some Theoretical Aspects of Destination Choice and Their Relevance to Production-Constrained Gravity Models. *Environment and Planning A: Economy and Space* **15**(8) 1121-1132.

Friedman, J., T. Hastie, R. Tibshirani. 2001. *The Elements of Statistical Learning* (Vol. 1). Berlin: Springer series in statistics, Springer.

Gao, H., J. Tang, H. Liu. 2012, 2012. *gSCorr: Modeling Geo-Social Correlations forNew Check-ins on Location-Based Social Networks.* Presentation, the 21st ACM international conference on Information and knowledge management, October 29 - November 2, Maui, HI.

Gaul, V. (2020). Location-based Services Market. Retrieved March 29, 2022, https://www.alliedmarketresearch.com/location-based-services-market#:~:text=The%20location%2Dbased%20services%20market%20size%20was%20valued%20at%20%2428.95,26.3%25%20from%202020%20to%202027.

Goh, K.-Y., C.-S. Heng, Z. Lin. 2013. Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. *Information Systems Research* **24**(1) 88-107.

Goodfellow, I., Y. Bengio, A. Courville. 2016. *Deep Learning*. The MIT Press, Cambridge, Massachusetts.

Grove, J. V. 2013. Foursquare gives out unsolicited tips on iPhone. Retrieved March 29, 2022, https://www.cnet.com/news/foursquare-gives-out-unsolicited-tips-on-iphone/.

Ho, Y.-J., S. Dewan, Y.-C. Ho. 2020. Distance and Local Competition in Mobile Geofencing. *Information Systems Research* **31**(4) 1421-1442.

Hochreiter, S., J. Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* **9**(8) 1735–1780.

Kilimci, Z. H., A. O. Akyuz, M. Uysal, S. Akyokus, M. O. Uysal, B. Atak Bulbul, M. A. Ekmis. 2019. An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain. *Complexity* **2019** 9067367.

Kim, W. G., J. Li, R. A. Brymer. 2016. The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management* **55** 41-51.

Kumar, V., A. Pansari. 2016. Competitive Advantage through Engagement. *Journal of Marketing Research* **53**(4) 497-514.

Lee, C., E. Ofek, T. J. Steenburgh. 2018. Personal and Social Usage: The Origins of Active Customers and Ways to Keep Them Engaged. *Management Science* **64**(6) 2473-2495.

Lee, G. M., L. Qiu, A. B. Whinston. 2016. A Friend Like Me: Modeling Network Formation in a Location-Based Social Network. *Journal of Management Information Systems* **33**(4) 1008-1033.

Li, X., D. Lord, Y. Zhang, Y. Xie. 2008. Predicting motor vehicle crashes using Support Vector Machine models. *Accident Analysis & Prevention* **40**(4) 1611-1618.

Li, Y., M. Steiner, L. Wang, Z.-L. Zhang, J. Bao. 2013. *Exploring venue popularity in foursquare.* Presentation, the 2013 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), April 14-19, Turin, Italy.

Lundberg, S. M., S.-I. Lee. 2017. *A unified approach to interpreting model predictions*. Presentation, the 31st International Conference on Neural Information Processing Systems, December 4-9, Long Beach, California, USA.

Pancras, J., S. Sriram, V. Kumar. 2012. Empirical Investigation of Retail Expansion and Cannibalization in a Dynamic Environment. *Management Science* **58**(11) 2001-2018.

Pinto, H., J. M. Almeida, M. A. Gonçalves. 2013. *Using early view patterns to predict the popularity of youtube videos*. Presenation, the sixth ACM international conference on Web search and data mining, Feburary 4-8, Rome, Italy.

Qiu, L., Z. Shi, A. B. Whinston. 2018. Learning from Your Friends' Check-Ins: An Empirical Study of Location-Based Social Networks. *Information Systems Research* **29**(4) 1044-1061.

Reunanen, J. 2003. Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research* **3** 1371–1382.

Rothenberger, S., D. Grewal, G. R. Iyer. 2008. Understanding the Role of Complaint Handling on Consumer Loyalty in Service Relationships. *Journal of Relationship Marketing* **7**(4) 359-376.

Rudin, C., D. Carlson. 2019. The Secrets of Machine Learning: Ten Things You Wish You Had Known Earlier to Be More Effective at Data Analysis. *Operations Research & Management Science in the Age of Analytics*, INFORMS, 44-72.

See-To, E., E. Ngai. 2018. Customer reviews for demand distribution and sales nowcasting: a big data approach. *Annals of Operations Research* **270** 415-431.

Sela, R. J., J. S. Simonoff. 2012. RE-EM trees: a data mining approach for longitudinal and clustered data. *Machine Learning* **86**(2) 169-207.

So, K. K. F., C. King, B. A. Sparks, Y. Wang. 2016. The role of customer engagement in building consumer loyalty to tourism brands. *Journal of Travel Research* **55**(1) 64-78.

Stiglic, G., P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, L. Cilar. 2020. Interpretability of machine learning ‐ based prediction models in healthcare. *WIREs Data Mining and Knowledge Discovery* **10**(5) e1379.

Sun, Z.-L., T.-M. Choi, K.-F. Au, Y. Yu. 2008. Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems* **46**(1) 411-419.

Szabó, G., B. Huberman. 2008. Predicting the Popularity of Online Content. *Communications of the ACM* **53**(8) 80-88.

Takenaka, T. 2020. Analysis and Prediction of Customer Behaviors for Restaurant Management. *Service Engineering for Gastronomic Sciences*, Springer, Singapore, 29-41.

Tatar, A., P. Antoniadis, M. Dias de Amorim, S. Fdida. 2014. From Popularity Prediction to Ranking Online News. *Social Network Analysis and Mining* **4** 174.

Vasconcelos, M., J. Almeida, M. Gonçalves. 2014. *What makes your opinion popular? predicting the popularity of micro-reviews in foursquare*. Presentation, the 29th Annual ACM Symposium on Applied Computing, March 24-28, Gyeongju, Republic of Korea.

Vasconcelos, M., J. M. Almeida, M. A. Gonçalves. 2015. Predicting the popularity of micro-reviews: A Foursquare case study. *Information Science* **325** 355-374.

Virrantaus, K., J. Markkula, A. Garmash, V. Terziyan, J. Veijalainen, A. Katanosov, H. Tirri. 2001. *Developing GIS-supported Location-Based Services*. Presentation, the 2nd International Conference on Web Information Systems Engineering, December 3-6, Kyoto, Japan.

Vivek, S. D., S. E. Beatty, R. M. Morgan. 2012. Customer Engagement: Exploring Customer Relationships Beyond Purchase. *Journal of Marketing Theory and Practice* **20**(2) 122-146.

Wang, L., R. Gopal, R. Shankar, J. Pancras. 2015. On the brink: Predicting business failure with mobile location-based checkins. *Decision Support Systems* **76** 3-13.

Xu, Y. 2012. Literature Review on Web Application Gamification and Analytics. *CSDL Technical Report* 11-05.

Zhang, D., J. Lin, Q. Peng, D. Wang, T. Yang, S. Sorooshian, . . . J. Zhuang. 2018. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology* **565** 720-736.

Zhang, J. Q., A. Dixit, R. Friedmann. 2010. Customer Loyalty and Lifetime Value: An Empirical Investigation of Consumer Packaged Goods. *Journal of Marketing Theory and Practice* **18**(2) 127-140.

Figures

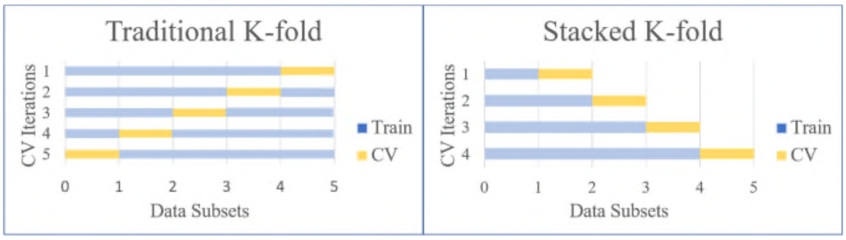

Figure 1. Cross-Validation

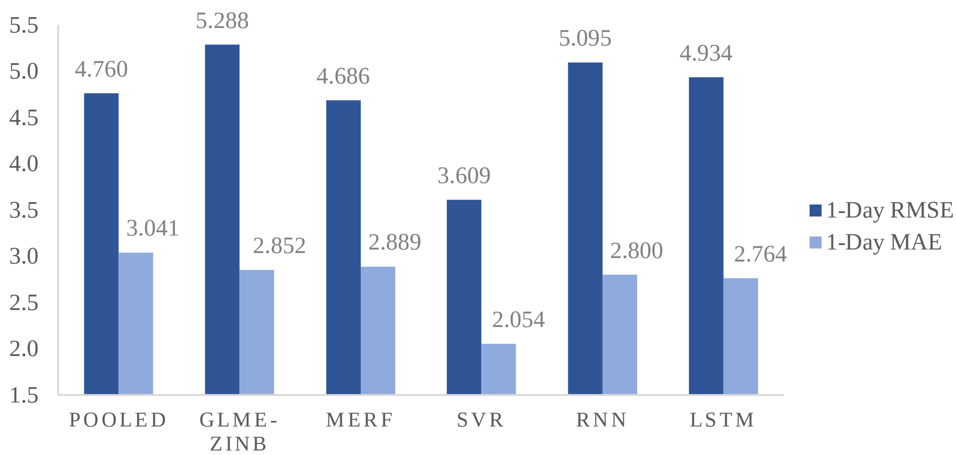

Figure 2. Prediction Accuracy Comparison (L-day=1)

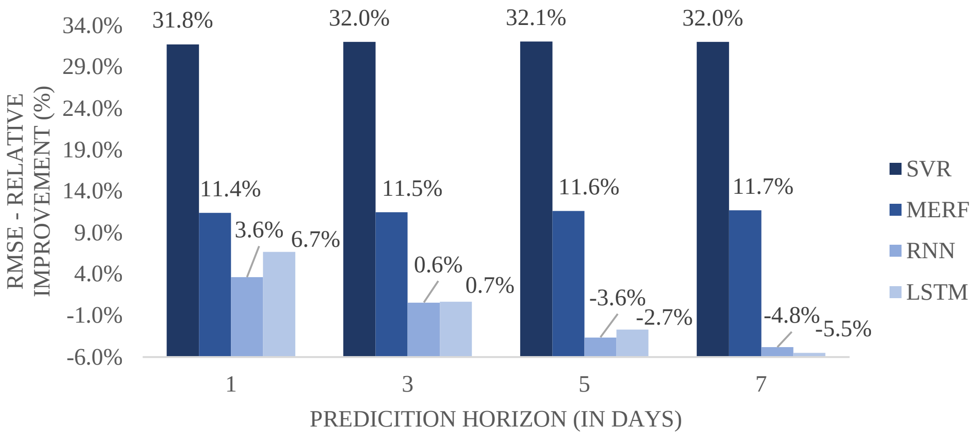

Figure 3. RMSE - Relative Forecast Improvement Over Prediction Horizon

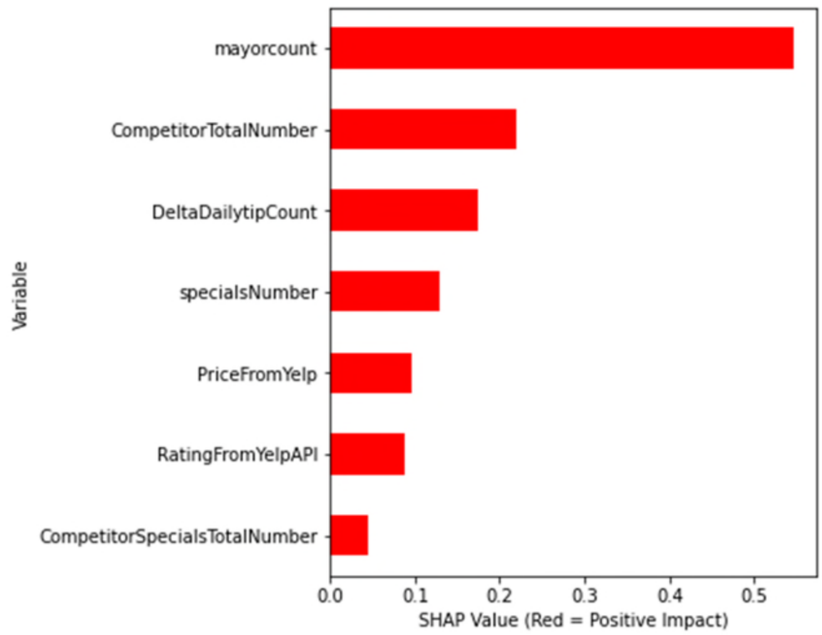

Figure 4. The SHAP Variable Importance Plot – Global Interpretability

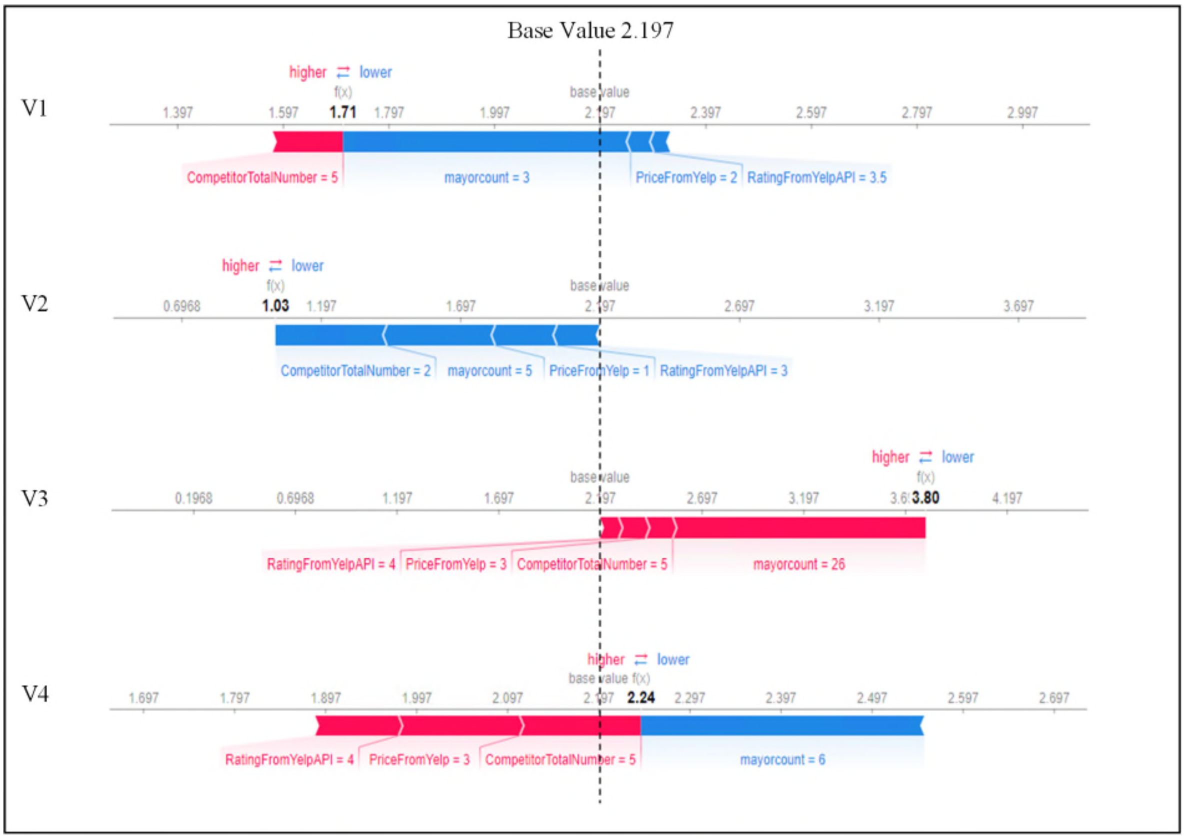

Figure 5. The SHAP Variable Importance Plot – Local Interpretability

Tables

Table 1. Summary Statistics

Variable Name	Min	Median	Mean	Max
Daily Check-in	0.00	2.00	3.64	296.00
Daily User	0.00	1.00	2.31	172.00
Daily Review	0.00	0.00	0.04	17.00
Mayor Check-in	0.00	7.00	10.75	60.00
Specials	0.00	0.00	0.10	3.00
CTN	0.00	5.00	4.07	5.00
CSTN	0.00	0.00	1.94	10.00
Yelp_Rating	1.00	3.50	3.56	4.50
Yelp_PriceRange	1.00	2.00	2.14	4.00

Table 2. Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Daily Check-in	1.00	0.93	0.21	0.22	0.09	0.12	0.05	0.02	0.06
(2) Daily User	0.93	1.00	0.20	0.16	0.08	0.12	0.05	0.02	0.11
(3) Daily Review	0.21	0.20	1.00	0.04	0.02	0.03	0.00	0.02	0.03
(4) Mayor Check-in	0.22	0.16	0.04	1.00	0.07	0.04	0.02	-0.03	0.00
(5) Specials	0.09	0.08	0.02	0.07	1.00	0.06	0.07	-0.04	0.03
(6) CTN	0.12	0.12	0.03	0.04	0.06	1.00	0.24	-0.05	0.14
(7) CSTN	0.05	0.05	0.00	0.02	0.07	0.24	1.00	0.01	0.07
(8) Yelp_Rating	0.02	0.02	0.02	-0.03	-0.04	-0.05	0.01	1.00	0.21
(9) Yelp_PriceRange	0.06	0.11	0.03	0.00	0.03	0.14	0.07	0.21	1.00

Table 3. Prediction Comparison for Six Algorithms (L-day=1)

	Out-of-Sample Error		In-Sample Error	
Model	RMSE	MAE	RMSE	MAE
Pooled	4.760	3.041	5.047	2.925
GLME-ZINB	5.288	2.852	5.613	2.875
SVR	3.609	2.058	3.436	1.743
MERF	4.686	2.889	4.050	2.243
RNN	5.095	2.800	5.548	2.908
LSTM	4.934	2.764	5.490	2.829

Table 4. Prediction Comparison for different lead days (L-day=1,3,5,7)

		Pooled	GLME-ZINB	MERF	SVR	RNN	LSTM
1-Day	RMSE	4.760	5.288	4.686	3.609	5.095	4.934
	MAE	3.041	2.852	2.889	2.054	2.800	2.764
3-Day	RMSE	4.755	5.288	4.681	3.594	5.257	5.251
	MAE	3.036	2.862	2.885	2.054	2.906	2.873
5-Day	RMSE	4.772	5.313	4.696	3.609	5.505	5.456
	MAE	3.042	2.876	2.892	2.061	3.041	2.974
7-Day	RMSE	4.774	5.321	4.699	3.617	5.577	5.615
	MAE	3.040	2.880	2.891	2.061	3.058	3.047

Table 5. Variable Selection Analysis – SVR

	1 Day		3 Day		5 Day		7 Day	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
All	3.436	1.743	3.444	1.749	3.441	1.741	3.443	1.734
Daily Review	3.458	1.750	3.467	1.756	3.463	1.748	3.466	1.741
Mayor Checkin	3.436	1.743	3.444	1.749	3.441	1.741	3.443	1.734
Specials	3.436	1.743	3.444	1.749	3.440	1.741	3.443	1.734
CTN	3.436	1.743	3.445	1.749	3.441	1.741	3.444	1.734
CSTN	3.436	1.743	3.445	1.749	3.441	1.741	3.444	1.734
Yelp_Rating	3.436	1.743	3.444	1.749	3.441	1.741	3.443	1.734
Yelp_PriceRange	3.436	1.743	3.444	1.749	3.440	1.741	3.443	1.734