

CARLSON SCHOOL
OF MANAGEMENT

UNIVERSITY OF MINNESOTA

Recruit Holdings: Restaurant Visitor Forecasting

Batool Fatima | Brady Engelke | Fazel | Jie Liu | Noah Becker | Patrick Seng



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Background

- Recruit Holdings is a promotional media conglomerate
- The Dining Services team in Japan intends to increase their value in the eyes of restaurant owners by marketing their visitor forecasting capabilities as a new product offering
- To establish forecasting credibility, they hope to accurately predict visitors for their current restaurant customers over a 39 day period

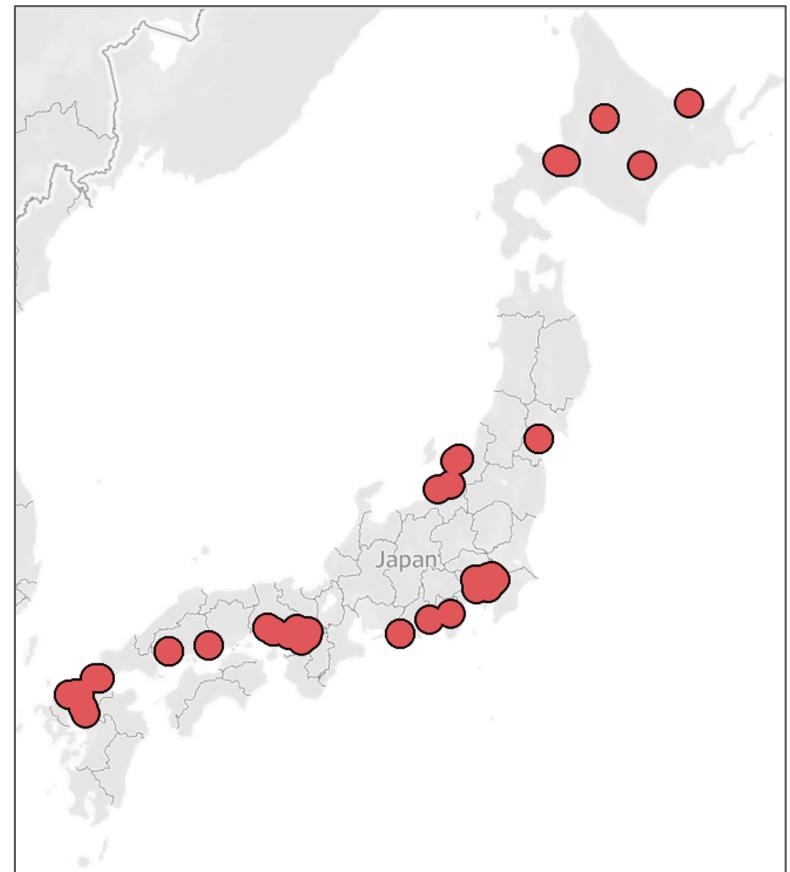


Business Impact

- We used 16 months of Recruit Holdings' restaurant data to make visitor predictions spanning 39 days and 829 restaurants across Japan
- Our best predictive model is a Light Gradient Boosting Regression Tree which accurately predicts visitors with an average error of ± 7 visitors per restaurant per day
- Impact
 - Establish added value to customers through demonstrated accurate visitor forecasts
 - Help Recruit Holdings' customers optimize staffing and supply chain efficiency

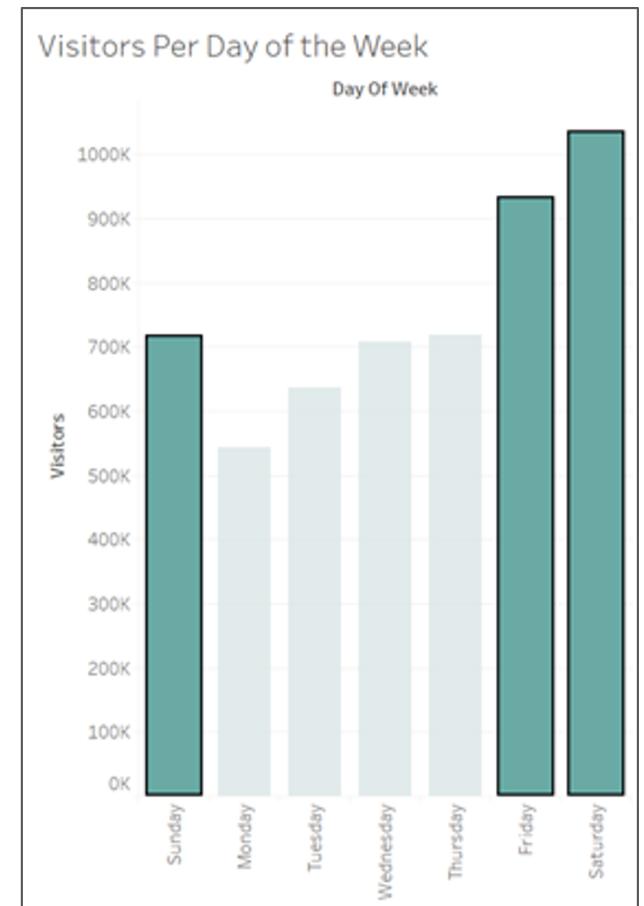
Recruit Holdings' Data

- AIR and HPG
- Restaurant Data:
 - Reservation Data - Reservations each restaurant had prior to each day of business
 - Store Info Data - Locations, genres of various restaurants
 - Visit Data - Actual totals of visitors per restaurant on any given day
 - Date info - National holidays in Japan
- 39 days, 829 restaurants, 14 genres, in 104 different areas



Data Preparation & Basic Features

- Merging of data sets
 - Ensure row for each restaurant each day
 - Drop unused data
- Basic Features:
 - Month of Year
 - Day of week, week of month
 - Day of, before, after holiday
 - Days restaurant has been in business



Feature Engineering

- Weather from Japan Meteorological Agency
 - Selected features: temperature, precipitation, sunlight, wind speed, humidity
 - Missing data treatment: K-NN
- 6, 7, 8, 9 Week Lookback
 - Avoid predictions upon predictions
 - Ensure data for most observations



Static Models

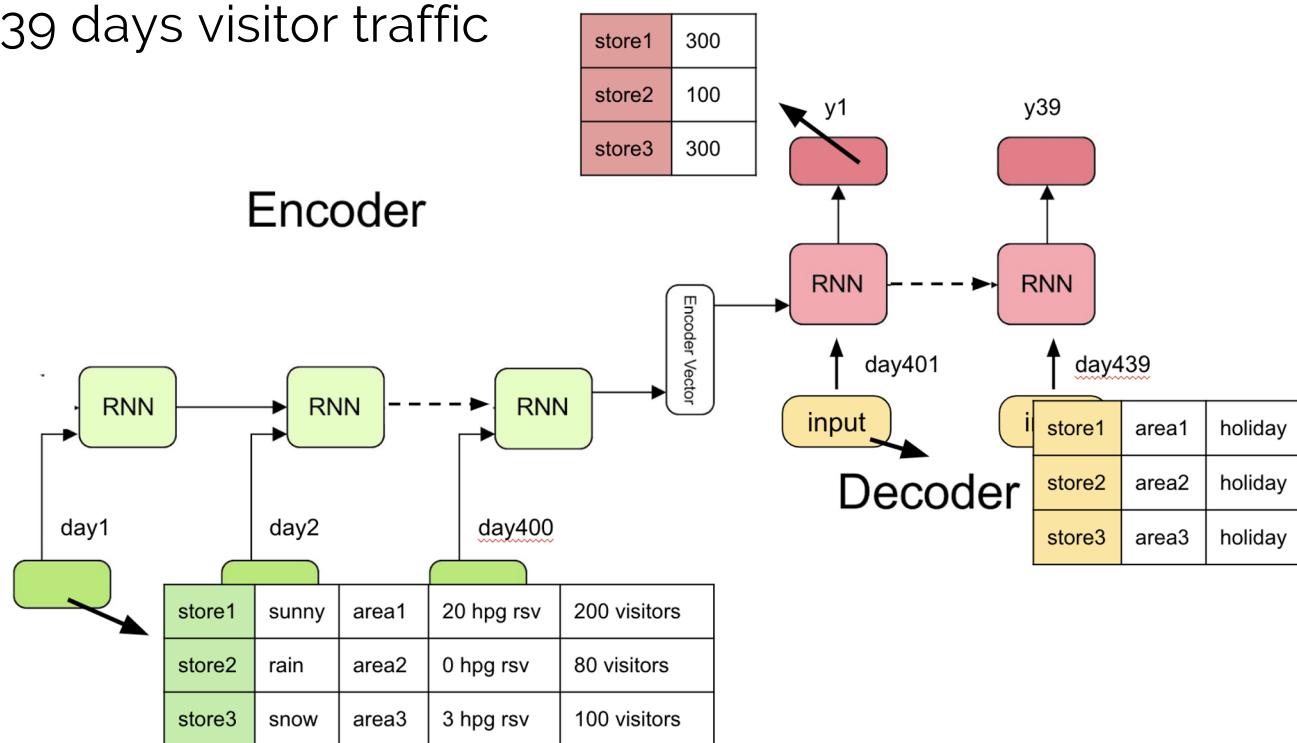
- Linear Regression
 - Interpretable coefficients for business users
- KNN
 - Quick training, slow, clunky testing
 - Performance affected by choice of distance metric
- Gradient Boosting Tree
 - Implies relative feature importances
 - Computationally costly
- Light Gradient Boosting Tree
 - Increased computational speed

RNN Model

- Seq2Seq can map arbitrary length sequences to other arbitrary length sequences
 - Sequence: Time Series
 - Arbitrarily length: We can look back and predict as many days as we want (as long as we have enough data to train the architecture)
- RNN/LSTM cells are implemented as encoder and decoder
 - Encoder memorizes and extracts the past pattern as states
 - Decoder takes encoder states and other input to generate prediction
- RNN/LSTM's ability to successfully learn on data with long range temporal dependencies makes it a natural choice for time series prediction

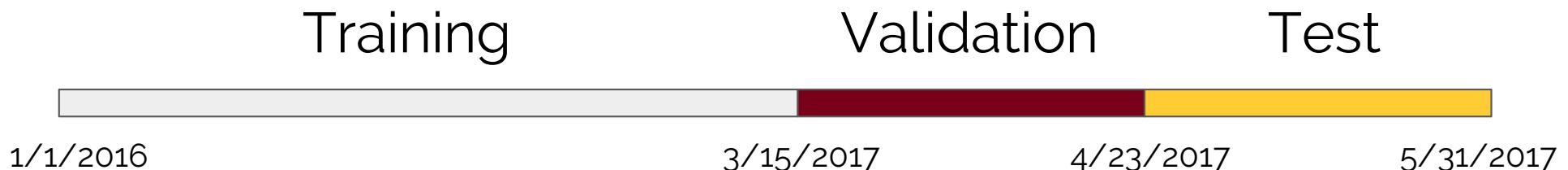
RNN Model

- We designed the architecture to take in past 400 days history to predict future 39 days visitor traffic



Validation & Evaluation Approach

- Strategic split of training, validation and testing data to mimic daily prediction of test data set

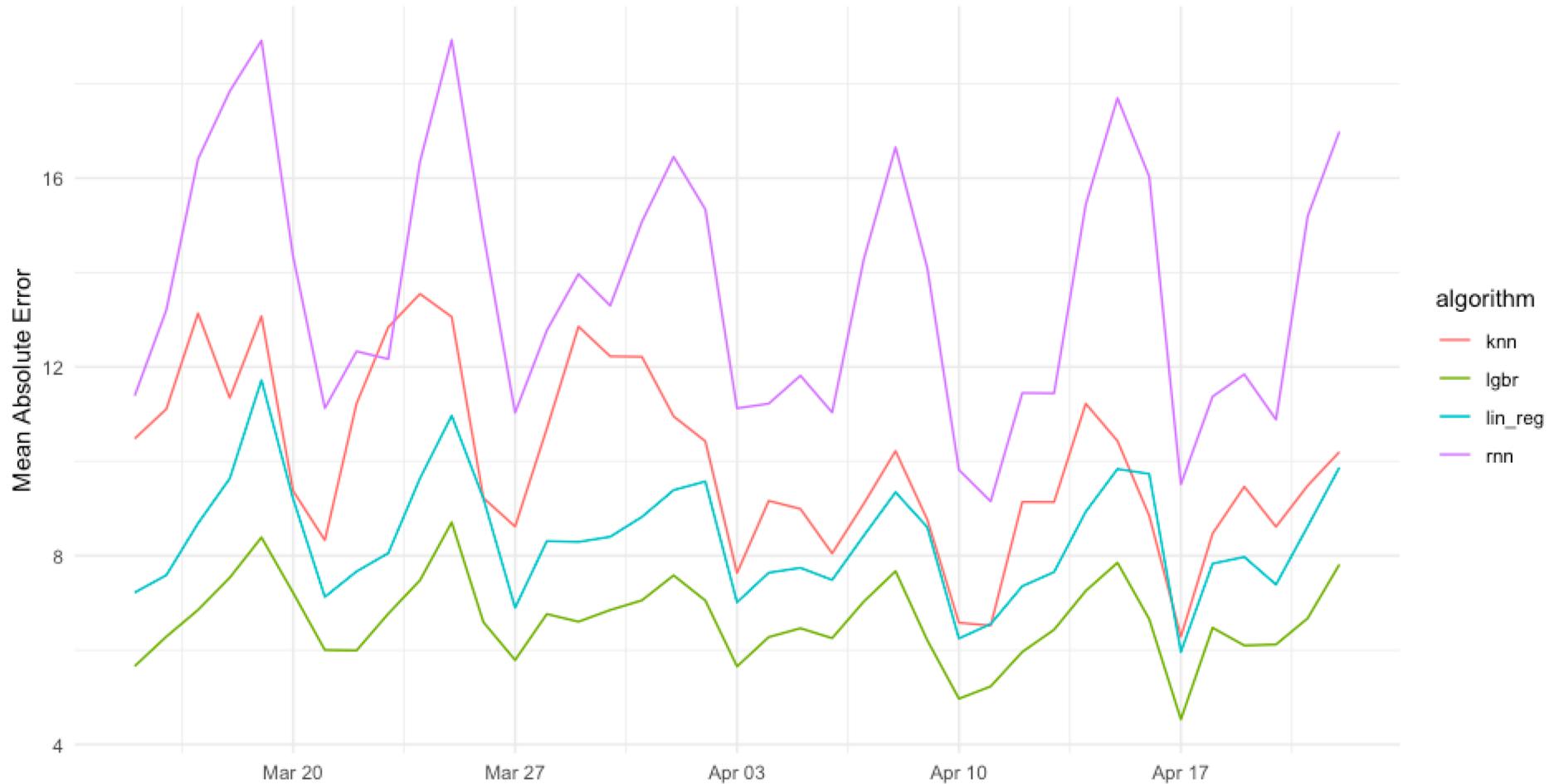


- Metrics used to evaluate models
 - MSE
 - RMSLE
 - R^2

Model Performance

Model	MSE	RMSLE	R ²
LGBR	135.39	0.536	0.613
Gradient Boost	143.981	0.541	0.588
Linear Regression	177.479	0.843	0.493
k-NN	272.347	0.982	0.221
RNN	403.608	1.230	0.201

Most Accurate Model: LGBR



Final Model: LGB

- Linear regression used as an appropriate “base case”
 - More accurate by 2 visitors per restaurant per day
- Feature importance scores provided

	features	importance
0	closed_flag	0.547238
1	lookback_visitors_avg_store	0.390829
2	reserve_visitors_air	0.027776
3	reserve_visitors_hpg	0.002308
4	dayofweek_5	0.002204
5	lookback_visitors_avg_genre	0.002078
6	holiday_flg	0.002025
7	open_days	0.001835
8	low_temperature	0.001452

Submission and Description	Private Score	Public Score	Use for Final Score
Submission_File_Fazel_closed.csv just now by Tabassum Fazel Recruit Restaurant LGB	0.58991	0.56789	<input type="checkbox"/>

Looking Forward

Prediction Improvement Opportunities:

- Additional genre details
- Additional features (ex. restaurant square ft.,)
- Improve RNN
- Refine look-back and weather functions

Implementation & Future Predictions:

- Update data regularly



Questions?