Professional Certificate in Artificial Intelligence and Machine Learning
UC Berkeley ExecEd 05/2025

Restaurant Success and Revenue Prediction Model Analysis

Marcus Lui

UC BERKELEY
EXECUTIVE EDUCATION

This is to confirm that

Marcus Lui

has successfully completed the

Professional Certificate in Machine Learning and Artificial Intelligence

May 2025

Tsu-Jae King Liu,
Dean, College of Engineering
Berkeley Engineering

Jennifer A. Chatman, Dean, Haas School of Business
BerkeleyHaas

What is the key measurement of success in any industry?

Factors that influence profit

Annual Revenue

Operational Efficiency*

Average Meal Price

Google Rating

Social Media Followers

Years in Business

Monthly Marketing Budget

*Cannot be considered for this study because it is usually proprietary information

The goals of this project are:

- To identify the best ML and AI models for restaurant success classification
- To identify the best ML and AI models for annual revenue prediction
- To provide information to help restaurants improve their success

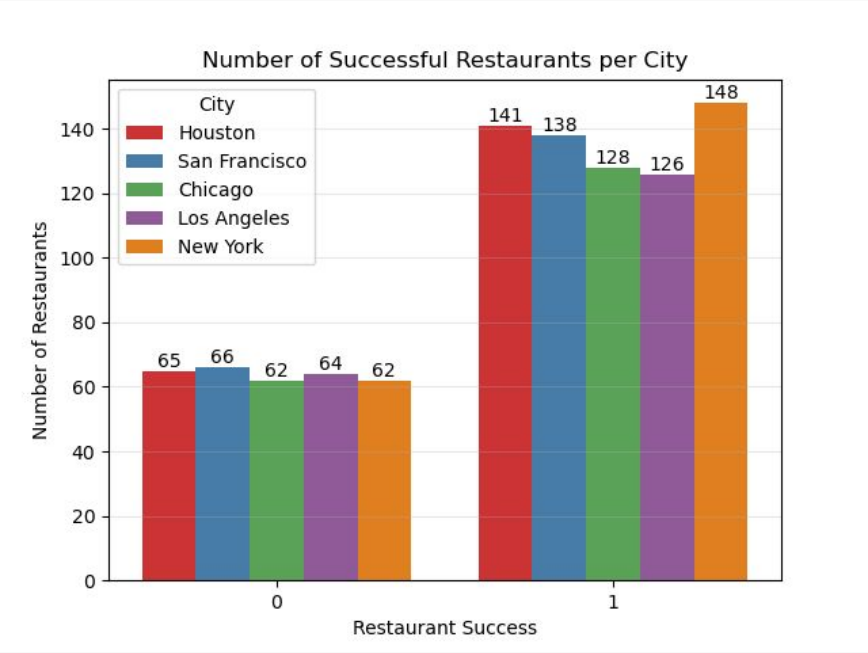
The *Data*

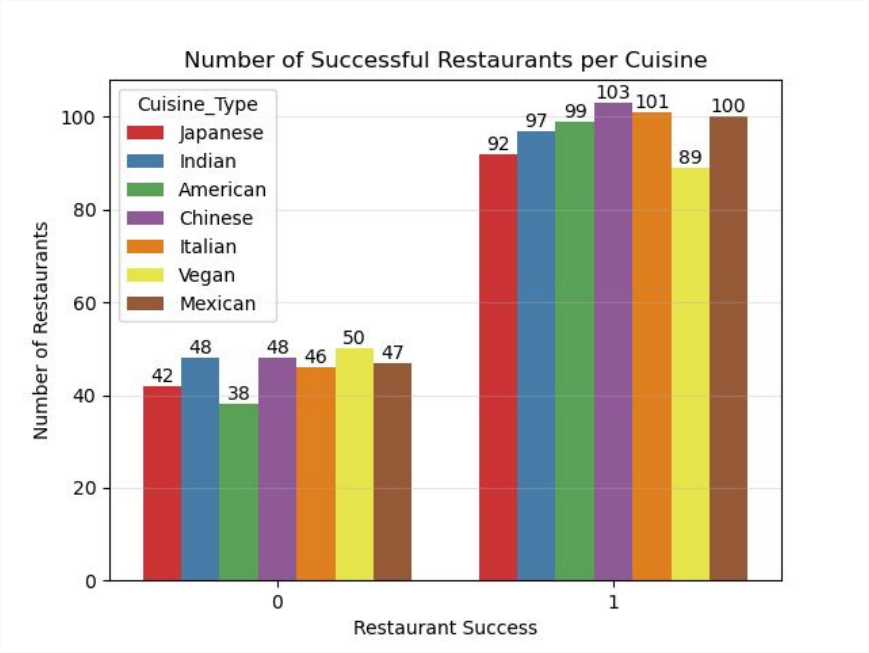
https://www.kaggle.com/datasets/liyangng/restaurant-success-prediction/data
1000 data points

- 9 Numerical Features
  - Average Meal Price
  - Seating Capacity
  - Years in Business
  - Google Rating
  - Social Media Followers
  - Weekend Reservations
  - Staff Count
  - Marketing Budget
  - Health Inspection Score
  - Annual Revenue

- 4 Categorical Features
  - City
  - Cuisine Type
  - Delivery Service
  - Success Label

The First Look

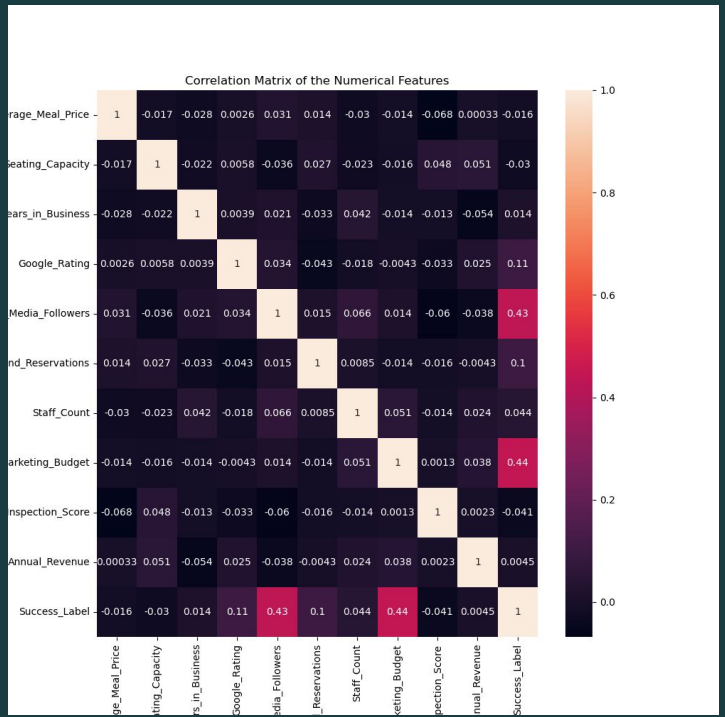


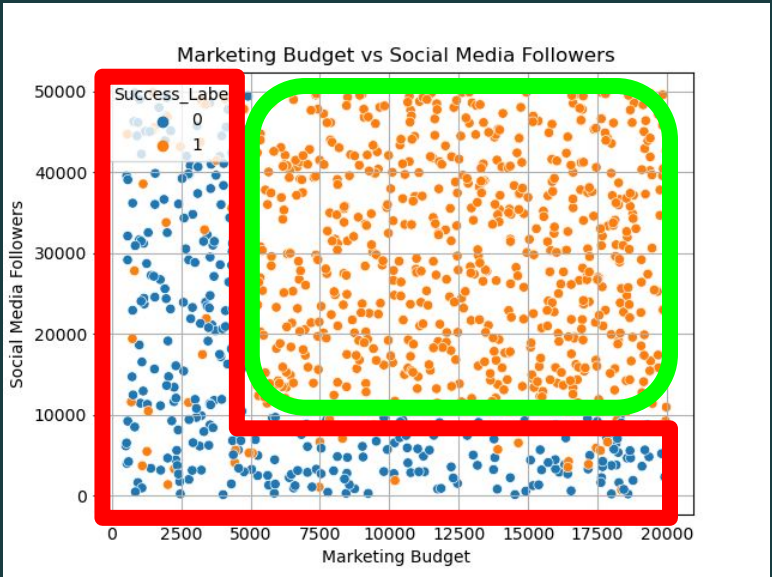


Approximately 2/3 of the restaurants were successful. Success was evenly distributed across cities and across cuisine type.

A Closer Look...

Social Media Followers and Market Budget had the highest correlation





There is a clear relationship between the features Social Media Followers and Marketing Budget

What model is best for classifying restaurant success?

Models for classifying restaurant success

Python via Anaconda and Jupyter Notebooks

Classification Models

- Dummy
- Logistic Regression
- K-Nearest Neighbors
- Decision Tree
- Random Forest

Classification Models

For restaurant success

Logistic Regression (LogReg)

A statistical model used for binary classification

K-Nearest Neighbors (KNN)

A simple, non-parametric algorithm used for classification

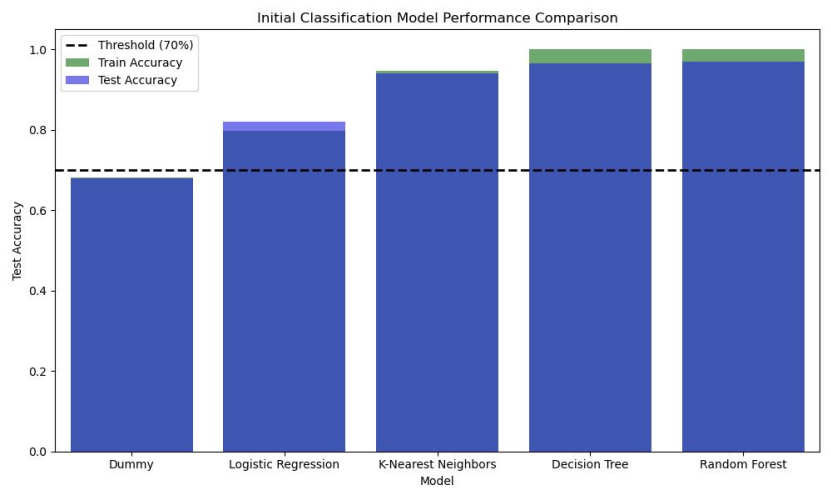
Decision Tree (DT)

A flowchart-like supervised learning algorithm used for classification

Random Forest (RF)

An ensemble learning method that builds multiple decision trees and combines their outputs to improve accuracy

Classification Model Analysis



4 of 5 models trained with over 80% accuracy

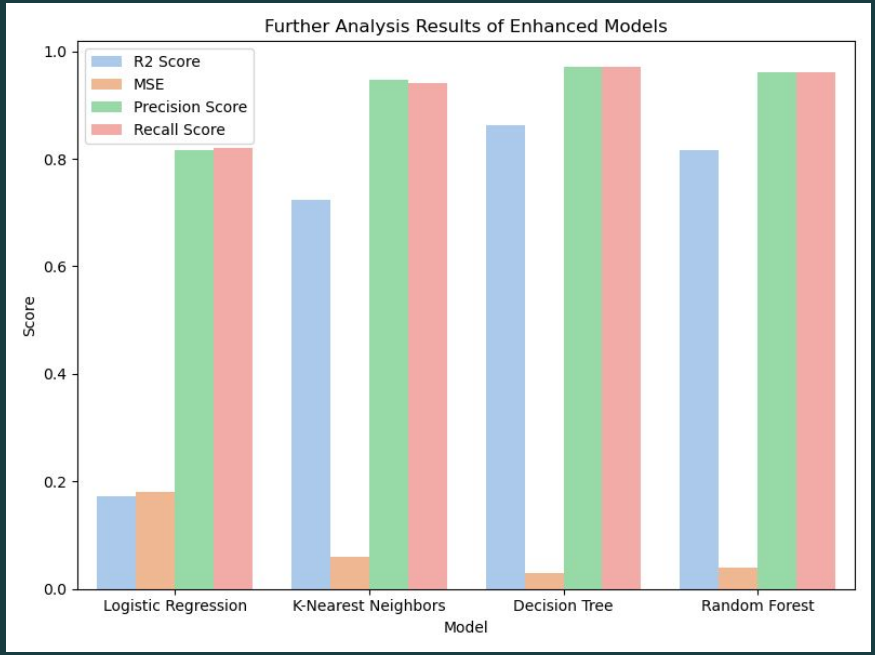
- Random Forest and Decision Tree
- GridSearchCV did not produce any notable improvements

A Deeper Look...

Decision Tree and Random Forest models trained the best

- High Recall Score
- High Precision Score
- High R^2 score
- Low Mean Squared Error

First Goal Completed



The goals of this project are:

- To identify the best ML and AI models for restaurant success classification
- To identify the best ML and AI models for annual revenue prediction
- To provide information to help restaurants improve their success

Models for predicting annual revenue

Python via Anaconda and Jupyter Notebooks

Regression Models

- Linear Regression
- Lasso
- Support Vector Regression
- Random Forest

Regression Models

For annual revenue prediction

Linear Regression (LinReg)

A simple supervised learning algorithm used to predict a continuous outcome

Lasso

A Linear Regression technique that includes L1 regularization to enhance model performance

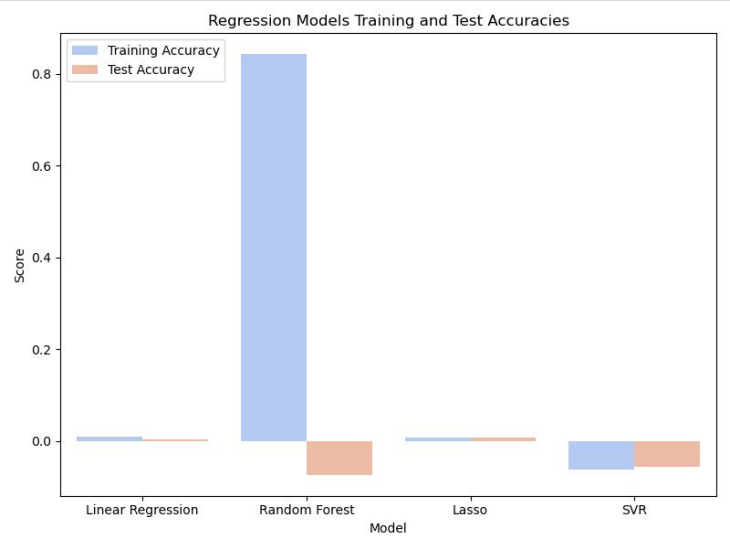
Support Vector Regression (SVR)

Predicts continuous values by finding a function that fits the data within a certain margin of tolerance, ignoring outliers

Random Forest (RF)

An ensemble machine learning method that uses multiple decision trees to predict a continuous output

Predictive Model Analysis



All models recorded a low testing accuracy

- RF recorded a high training accuracy
- Applied GridSearchCV
  - Results were indistinguishable

No models were good at predicting revenue

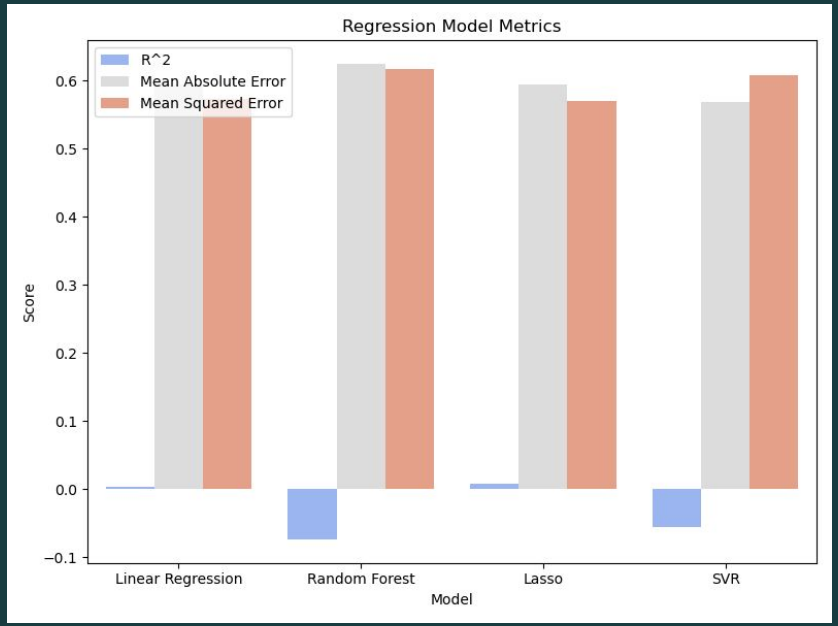
- Potential problems could be overfitting or bad data

A Deeper Look...

No models were able to predict annual revenue with high accuracy

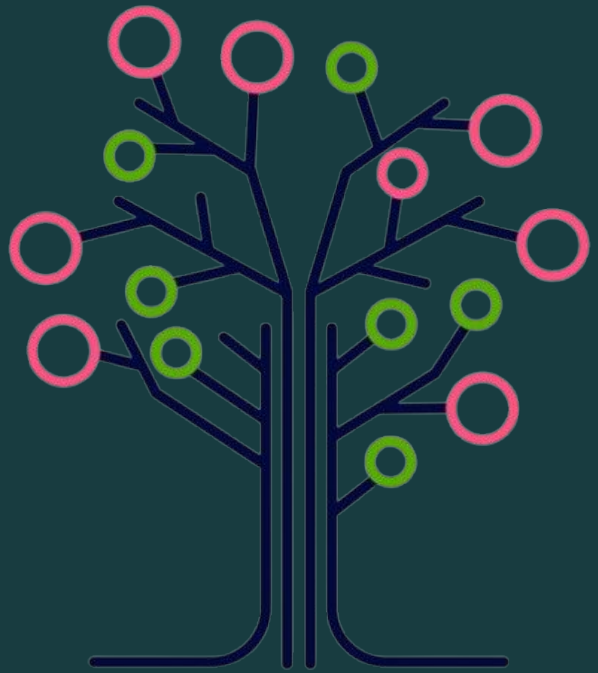
- High Mean Absolute Error
- High Mean Squared Error
- Low R^2 score

Second goal completed... sort of



The goals of this project are:

- To identify the best ML and AI models for restaurant success classification
- To identify the best ML and AI models for annual revenue prediction
- To provide information to help restaurants improve their success



Recommendations for improving success

- Increase monthly marketing budget to a range between $5k and $20k
- Increase social media following across Instagram and TikTok to over 10,000

To make a better prediction model...

Models require higher quality data

- Foot traffic metrics and table turnover rate
- Customer satisfaction
- Repeat customer rate
- Most effective means of marketing

Thank you

https://github.com/MarcusLui9/Restaurant_Success_and_Revenue_Prediction
https://www.linkedin.com/in/marcus-chi-kin-lui/