



CAL POLY POMONA

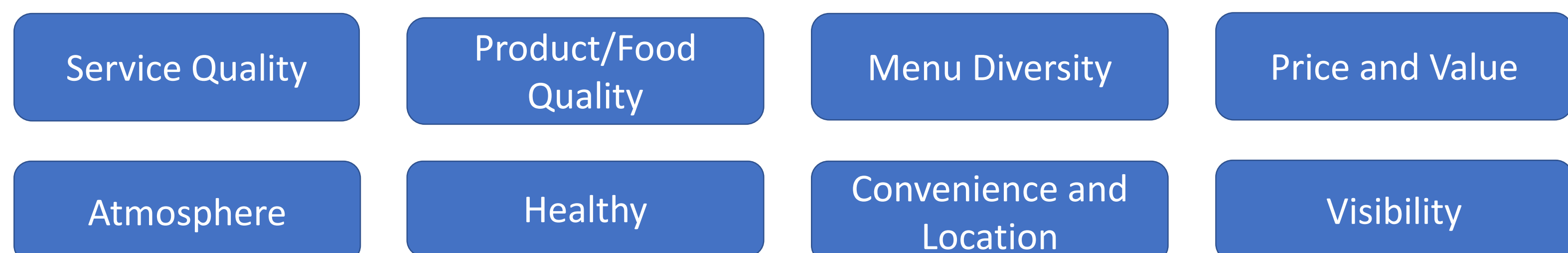
A Recommender System for Restaurants Based on Review Factors, Reviewer Sentiment, and Restaurant Culture Category

Ishmene Andrei Cabral, Computer Information Systems
Mentor: Dr. Sonya Zhang
Kellogg Honors College Capstone Project

Abstract

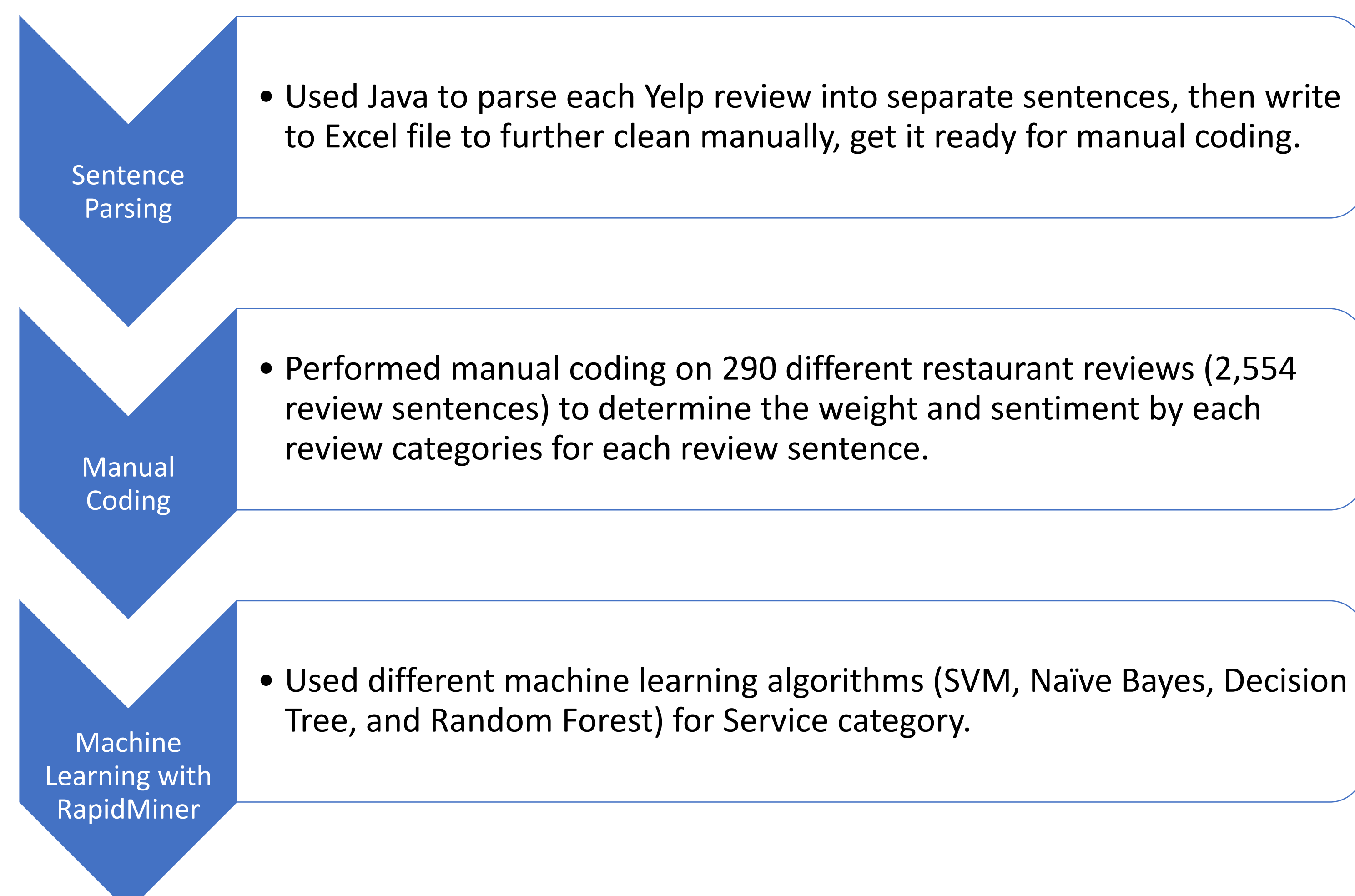
As consumers are becoming more reliant on online reviews for making everyday decisions such as which restaurant to dine at, local businesses owners also use these consumer reviews to increase visibility as well as to get insights for business improvement. At the same time, the large volume of the reviews and limited resources can make it challenging to extract, analyze and drive useful knowledge. In this project, we aim to automate the categorization of the restaurant reviews by different review factors such as service quality and staff attitude, product quality and hygiene, adventurous menu, price and value, atmosphere and appearance, healthy food, convenience and location, and visibility. Using customer reviews and restaurant data from a recent Yelp dataset, we performed procedures including sentence parsing, manual coding to determine the weight and sentiment of each factor in the review, and machine learning algorithms in efforts to find the best prediction models for each factor. Eventually we hope to use the models to explore if there is a variety of consumer perspectives among different cultural categories of restaurants.

Categories for Restaurant Reviews



- ★ **Service Quality:** The standards, consistent quality, courtesy of the service; friendliness, knowledge, willingness to help, communication, competency, and attentiveness of the restaurant staff.
- ★ **Product/Food Quality:** The quality, portions, tastiness, temperature, presentation, preparation consistency and non-greasiness of food.
- ★ **Menu Diversity:** The menu variety, or availability of menu items, dishes or beverages liked, local dishes, and health food choice.
- ★ **Price and Value:** The prices and value of food.
- ★ **Atmosphere:** The atmosphere or ambience in the restaurant.
- ★ **Healthy:** The availability of healthy food and nutritious food.
- ★ **Convenience and Location:** The location, crowd level, and operating hours of the restaurant.
- ★ **Visibility:** The visibility of food preparation area.

Procedure



Machine Learning with RapidMiner

Used different machine learning algorithms:

- ➔ **SVM:** Also known as Support Vector Machine, can be used for classification, regression, or other tasks like outliers detection.
- ➔ **Naive Bayes:** a high-bias, low-variance classifier that can build a good model even with a small data set. It is typically used in cases that involve text categorization, including spam detection, sentiment analysis, and recommender systems.
- ➔ **Decision Tree:** a tree like collection of nodes intended to create a decision on values affiliation to a class or an estimate of a numerical target value. Each node represents a splitting rule for one specific Attribute. Can be used for classification and regression.
- ➔ **Random Forest:** an ensemble of a certain number of random trees, specified by the *number of trees* parameter. These trees are created/trained on bootstrapped sub-sets of the ExampleSet provided at the Input Port. For classification the rule is separating values belonging to different classes, while for regression it separates them in order to reduce the error made by the estimation.

Results

Service Factor Performance	SVM	Naive Bayes	Decision Tree	Random Forest
True 1 Pred. 1	64.11%	78.85%	64.79%	59.59%
True 0 Pred. 0	86.01%	89.14%	96.11%	98.53%
F-measure	63.12%	92.23%	70.24%	70.98%

In our results, Naive Bayes yield the best results when analyzing the Service Quality factor. In addition to using machine learning algorithms, we used dimension deduction which filters stop words using customized a dictionary creating a example set to cluster words into coherent topics. For example, 'waitress' and 'bartender'. We also used feature selection (pruning), and 20 fold cross validation. Using these models, in the future we hope to explore if there is a variety of consumer perspectives among different cultural categories of restaurants.

References

- RapidMiner. "Random Forest (Concurrency)." *RapidMiner Documentation*, docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/parallel_random_forest.html
- RapidMiner. "Decision Tree (Concurrency)." *RapidMiner Documentation*, docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/parallel_decision_tree.html
- RapidMiner. "Naive Bayes (RapidMiner Studio Core)." *RapidMiner Documentation*, docs.rapidminer.com/latest/studio/operators/modeling/predictive/bayesian/naive_bayes.html
- Salehan, M., Zhang, S., Aghakhani, N. 2018. "A recommender System for Restaurant Reviews Based on Consumer Segment", Proceedings of the 25th Americas Conference on Information Systems (AMCIS)