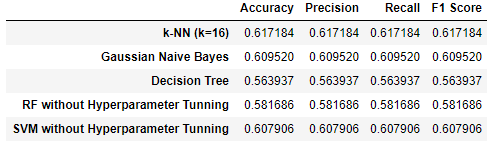
**Reflection and Conclusion:**

1. **What is the relationship between Business location and local catering industries?**

For this question, we can define the local catering industries by the star of the restaurant. From the analysis, we can see that most restaurants have 3.5 to 4.0 stars, and the majority number of reviews is lower than 500. For the restaurant categories, Mexican, Italian, and Chinese are the top three popular.

From the chart, we can see that the relationship between the stars and other factors that quite different. The most significant part is the latitude, which has a coefficient of 0.0196. That means that the high latitude will result in the high stars. What is more, with the increasing of the stars, all category countries have a negative relationship, which means that the high star restaurant has strictly standard.

We get useful information by using the bar chart and logical regression analysis. We tried to find the best model and compared K-Nearest Neighbors, Gaussian Naive Bayes, Decision Tree, Random Forest, and SVM.



We found that KNN is the best model for this dataset with an F1 score to 61% and 61,7% accuracy, which means this model can recognize the high star rating restaurants most precisely. What is more, the Naïve Bayes and SVM without Hyperparameter Tuning have the same performance with similar accuracy. Therefore, KNN better fits our goal.

1. **What factors impact on customer satisfaction for a specific restaurant?**

From the analysis result, Business parking, restaurant reservation, restaurant good for groups, business accepts credit cards, take out service have a positive effect on customer satisfaction. If the restaurant can provide this, it will improve customer satisfaction. However, a parking garage, parking validated, parking valet, or near the street, provide delivery and have free WiFi not affect customer satisfaction. We get useful information by using the bar chart and logical regression analysis. We tried to find the best model and compared SVM and RF. 手机屏幕截图

描述已自动生成

We found that random forests with Hyperparameter Tuning are the best model for this dataset with an F1 score high up to 86%. However, SVM\_rbf performed with the best precision, which means this model can recognize the high star rating restaurants most precisely. We need to choose the model according to performance as well as our objective. Our question is to figure out factors impacting on customer satisfaction (star rating). We need to analyze the features of delicious restaurants. Therefore, SVM\_rbf better fits our goal.

1. **What is the most important characteristic of elite users?**

For this question, we can find that most elite users have less than two years, and the number of reviews is also less than 2000 reviews. For elite users, the majorite factor is review count, useful, review level.

We compared four models including KNN, Gaussian Naive Bayes，Bernoulli Naive Bayes and decision tree.手机屏幕截图

描述已自动生成

We found that Decision Tree has the highest accuracy and F1 score, while Bernoulli naive Bayes has the highest Recall and Gaussian naive Bayes has the highest precision. Based on the data we predict, we found that the number of user reviews will affect the number of elite years. Therefore, we suggest the restaurant try to encourage customers to use yelp and give positive comments.

Based on the data we predict, we found that the number of user reviews will affect the number of elite years. Therefore, we suggest the restaurant try to encourage customers to use yelp and give positive comments. For example, every time you go to a restaurant and use yelp reviews, you can get points. Accumulate points based on the amount of each consumption. Then prepare the corresponding points to exchange for rewards. The reward can be a coupon. The most important thing is that a restaurant with high-quality food and good service can attract customers to give positive comments.