Module Tutor: Iftikhar Afridi

Decoding Customer Sentiment on Restaurant Reviews using R

COM692 Data Analytics 2024/2025

Sharad Thing

B00864068

Introduction to Domain

The Restaurant Industry: A Sector Under Pressure

❖ **2025 UK Restaurant Bankruptcies**

- Total: 1,932
- Average Closures: 5+ daily
- Increase: 45% from 2022

High-profile closures include Le Gavroche and Pollen Street Social

❖ **Solution:**

- ML model to classify sentiments (Positive/Negative) and aspects (Food/Service/Price).

❖ **Dataset:**

- Source: "Restaurant_reviews.csv" (Kaggle) containing (Restaurants, Reviews, ratings, timestamps)

Objectives of the Analysis

Objectives: Clear Goals

1.Primary Goal:

- Automatically classify reviews into sentiment (Positive/Negative) and classify sentiments into (Food/Service/Price/Other).

2.Key Tasks:

- Load & Clean Data
- Understand the dataset
- Descriptive & EDA

- Machine Learning - Model 1 – Train a classifier to detect overall sentiments (Positive/Negative) via text Reviews.
- Machine Learning - Model 2 – Classify positive/negative reviews by aspects like food, service, and price.

- Optimisation & Results – Fine-tune models and interpret results with metrics

Fig. 1: Source: Online source

Data Cleaning and Enhancing for ML

Data Cleaning:

- Missing values, irrelevant columns (e.g., Pictures).
- Removed unnecessary columns. Dropped irrelevant features like Pictures.
- Converted date-time formats for temporal analysis.
- Handled missing(NA) values by removing them
- Added “Sentiments”, “Sentiments Description” columns for 2 layered ML

```{r}
#Remove unnecessary column
reviews <- reviews %>% select(-`7514`)
colnames(reviews)
# Remove rows with missing values and drop Pictures
reviews <- reviews %>%
  drop_na() %>%
  select(-Pictures)

```

Text-Processing Steps before Machine Learning:

1. Convert to Lowercase
2. Remove Numbers
3. Remove Punctuation & Special Characters
4. Remove Stop words
5. Strip Whitespace

```{r}
# Clean text data through multiple steps:

# Step 1: Convert to lowercase
corpus_clean <- tm_map(corpus, content_transformer(tolower))
# Step 2: Remove numbers
corpus_clean <- tm_map(corpus_clean, removeNumbers)
# Step 3: Remove punctuation
corpus_clean <- tm_map(corpus_clean, removePunctuation)
corpus_clean <- tm_map(corpus_clean, content_transformer(function(x) gsub("’|‘|`|'", "", x)))
# Step 4: Remove common stopwords like "the", "is", etc.
corpus_clean <- tm_map(corpus_clean, removeWords, stopwords("english"))
# Step 5: Remove extra spaces
corpus_clean <- tm_map(corpus_clean, stripWhitespace)
# Step 6: Apply stemming (e.g., "amazing", "amazingly" → "amaz")
#corpus_clean <- tm_map(corpus_clean, stemDocument)

```

Fig. 2: Source: R Script

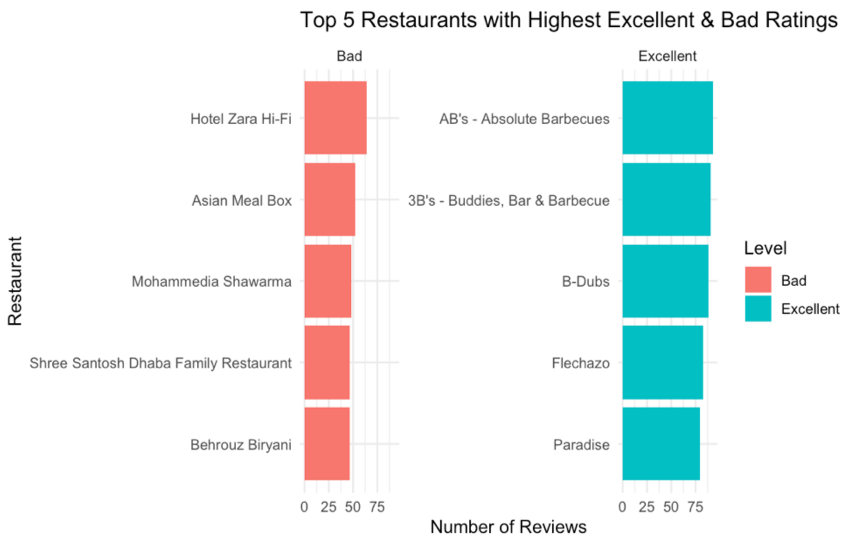
Exploratory Data Analysis (EDA)

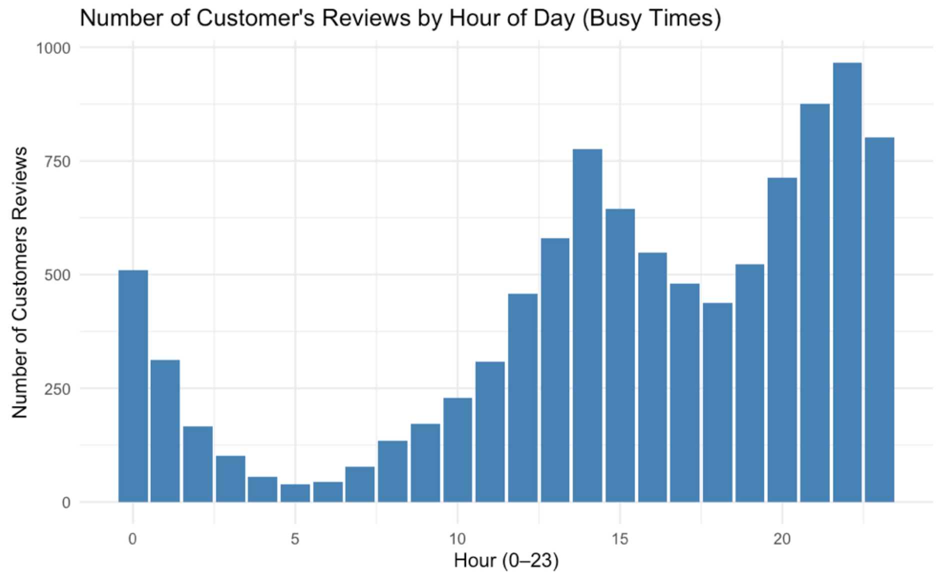
❖ Exploratory Data Analysis (EDA):

- Visualised rating distributions and busiest review hours

- Analysed sentiment vs. rating using bar and pie charts

- Identified top restaurants and most loyal reviewers





Machine Learning Approach using R

Sentiment Labelling

- ❖ **Feature Extraction:**
  • Built a Document-Term Matrix (DTM) and removed sparse terms
  • Extracted the most frequent keywords used across all reviews

- ❖ **Initial Sentiment Labelling:**
  • Used rule-based labelling with keywords and review text to tag sentiments (Positive/Negative)
  • Saved the partially labelled dataset for machine learning training

❖ **Sentiments Based on Keywords:**

1. Positive keywords: "polite", "wonderful", "good", "excellent", "awesome", "quick".
2. Negative keywords: "bad", "worst", "didn't", "poor", "rude", "tasteless".

Ulster University

Machine Learning Approach

- **Machine Learning Model (Binary Classification)**
  - Split labelled reviews into training/testing sets (80/20 split).

- **Modelling (1st Model):**
  - Trained a Support Vector Machine (SVM) using caret
  - Sentiment analysis as Positive or Negative Reviews

- **Modelling (2nd Model):**
  - Trained another **SVM model** on the result obtained via 1st Model.
  - Sentiment Classification: food, price, service and other

```r
```{r}
#1library(e1071)

#nb_model <- naiveBayes(Label ~ ., data = train_data)


# Set up training control
ctrl <- trainControl(method = "cv", number = 10 ,verboseIter = TRUE)

# Train model using caret with, say, Random Forest (rf)
model <- train(
  Sentiments ~ .,
  data = train_data,
  method = "svmLinear" ,
  trControl = ctrl
)

# Predict on test data
predictions <- predict(model, newdata = test_data)


```
```

Optimization, Evaluation and Metrics

Model Optimization

Model Optimisation (Slide)

•**Initial model**: Naive Bayes using e1071

- Accuracy: ~65%
- Train/Test split: 70/30
- Performance affected by small dataset & simplistic assumptions

•**Switched to**: SVM (Support Vector Machine) with caret

- Used **10-fold cross-validation**
- Better suited for high-dimensional text data
- Achieved **95% accuracy and more reliable predictions**

•**Final choice**: SVM Model

- Used for both:
  1. sentiment polarity
  2. aspect-based classification

```
# ---------------------
# 12. Model Training & Evaluation
# ---------------------
```

```
```{r}
#library(e1071)

#nb_model <- naiveBayes(Label ~ ., data = train_data)


# Set up training control
ctrl <- trainControl(method = "cv", number = 10, verboseIter
```
```

Metrics

Model 1: Sentiments

- Accuracy: 95.09%
- Kappa: 0.81 (strong agreement)
- Sensitivity (Recall for Positive): 97.85%
- Specificity (Recall for Negative): 80.48%
- Precision (Positive): 96.38%
- Balanced Accuracy: 89.16%

Model 2: Sentiment Classification

- Accuracy: 87.0%
- Kappa: 0.71 (strong agreement)
- Specificity (Recall): 89%
- Precision (Positive): 86.38%
- Balanced Accuracy: 89.16%

Model Accuracy

```
Confusion Matrix and Statistics

                 Reference
Prediction         Positive Review Negative Review
  Positive Review            1091              41
  Negative Review              24             169

               Accuracy : 0.9509
                 95% CI : (0.9379, 0.9619)
    No Information Rate : 0.8415
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.8098

 Mcnemar's Test P-Value : 0.04719

            Sensitivity : 0.9785
            Specificity : 0.8048
         Pos Pred Value : 0.9638
         Neg Pred Value : 0.8756
             Prevalence : 0.8415
         Detection Rate : 0.8234
   Detection Prevalence : 0.8543
      Balanced Accuracy : 0.8916

       'Positive' Class : Positive Review
```

Sentiment Model 1 (Positive/Negative)

```
Confusion Matrix and Statistics

              Reference
Prediction     Good Food Good Service Affordable Bad Food Bad Service Overpriced Other
  Good Food         1182            9          1       61           2          1     1
  Good Service        14           26          0        1           2          0     0
  Affordable           8            0          1        1           0          0     0
  Bad Food            73            2          1      334          12          1     1
  Bad Service          7            1          0       14           8          0     0
  Overpriced           1            0          0        4           1          7     0
  Other                1            0          0        0           0          0     0

Overall Statistics

               Accuracy : 0.8763
                 95% CI : (0.8601, 0.8912)
    No Information Rate : 0.7233
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7137
```

Sentiment Model 2 (Classification)

Results/
Visualisations

Ulster University

Results

	Reviewer	Restaurant	Rating	Review	Sentiments
1	Vicky	Sardarji's Chaats & More	1	I received wrong order..i tried to contact with restaur…	Negative Review
2	Sam	NorFest – The Dhaba	1	Was served very bad quality of chicken, even for the s…	Negative Review
3	Sam	Tiki Shack	1	We visited the place on Thursday night after calling u…	Negative Review
4	Priyanka	Karachi Cafe	2	The retro theme ambience was classic. But I didnt like…	Negative Review
5	Suresh	La La Land – Bar & Kitchen	2	We went for lunch buffet. Food taste is average and li…	Negative Review
6	Santosh	Owm Nom Nom	1	Worst experience ever. Food was spoiled and the sam…	Negative Review
7	Priyanka	Asian Meal Box	1	not good.	Negative Review
8	Suresh	Being Hungry	1	very bad.. ordered special biryani,but it was mixed wi…	Negative Review
9	Priyanka	Eat India Company	1	No [illegible] Wouldn't recommend to anyone. The…	Negative Review
10	Suresh	Eat India Company	1	Too much of crowd and no one will attend you for qui…	Negative Review

Sentiment Model 1 (Positive/Negative)

	Reviewer	Restaurant	Rating	Review	Sentiments
1	Manojkumar D Nambisan	Shah Ghouse Hotel & Restaurant	3.0	Food is decent and tastes palatable, not the best, tho…	Positive Review
2	Ankita	Hyper Local	4.0	We liked the chhole bhature as it was not at all oily. T…	Positive Review
3	Vedant Killa	Barbeque Nation	5.0	Excellent service and food buffet! This place has reall…	Positive Review
4	Ankita	The Lal Street – Bar Exchange	4.0	A good place to hangout. They have indoor as well as…	Positive Review
5	Manojkumar D Nambisan	The Lal Street – Bar Exchange	4.0	Good food options and decent service. Is expensive a…	Positive Review
6	Kiran	10 Downing Street	4.0	The place is decent and not heavily crowded. The am…	Positive Review
7	Manojkumar D Nambisan	10 Downing Street	3.0	Food is average, service is poor. Can get crowded. Th…	Positive Review
8	Ankita	Jonathan's Kitchen – Holiday Inn Express & Suites	4.0	A quick dinner for last Sunday of 2018 landed us in J…	Positive Review

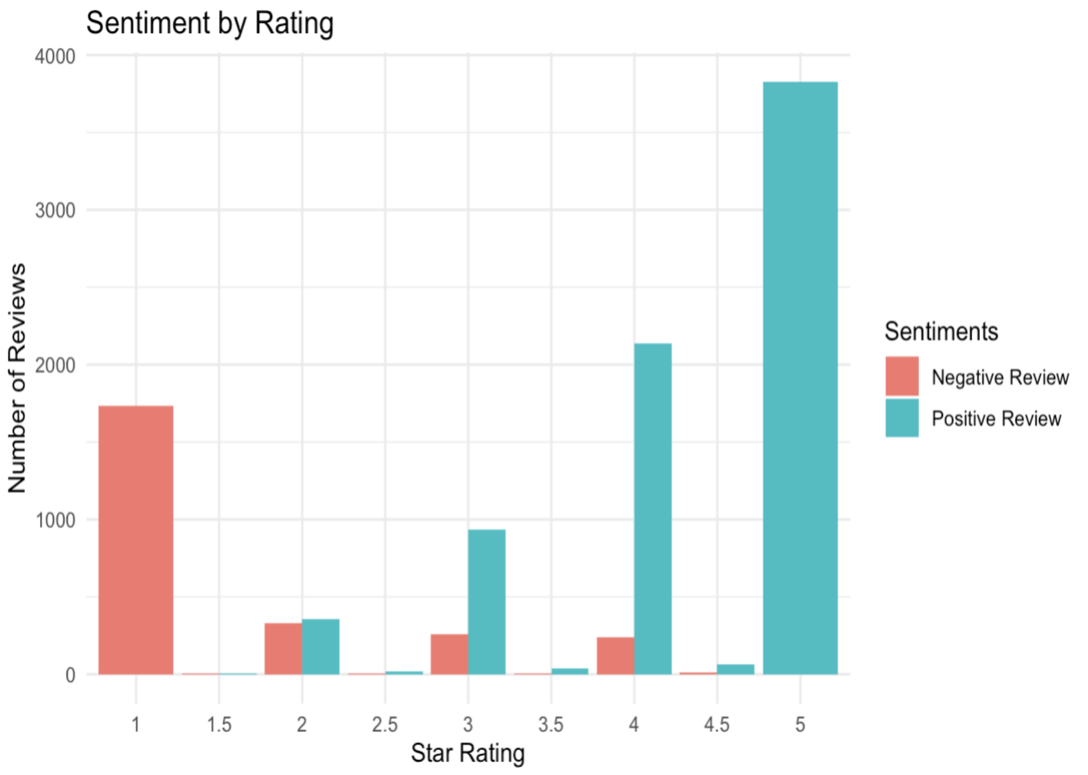
Results

	Reviewer	Restaurant	Rating	Review	Sentiments	Sentiment_Description
1	Vicky	Sardarji's Chaats & More	1	I received wrong order..i tried to contact with restaur…	Negative Review	Bad Service
2	Sam	NorFest – The Dhaba	1	Was served very bad quality of chicken, even for the s…	Negative Review	Bad Food
3	Sam	Tiki Shack	1	We visited the place on Thursday night after calling u…	Negative Review	Bad Food
4	Priyanka	Karachi Cafe	2	The retro theme ambience was classic. But I didnt like…	Negative Review	Bad Food
5	Suresh	La La Land – Bar & Kitchen	2	We went for lunch buffet. Food taste is average and li…	Negative Review	Bad Food
6	Santosh	Owm Nom Nom	1	Worst experience ever. Food was spoiled and the sam…	Negative Review	Bad Food
7	Priyanka	Asian Meal Box	1	not good.	Negative Review	NA
8	Suresh	Being Hungry	1	very bad.. ordered special biryani,but it was mixed wi…	Negative Review	Bad Food
9	Priyanka	Eat India Company	1	No proper food. Wouldn't recommend to anyone. The…	Negative Review	Bad Food
10	Suresh	Eat India Company	1	Too much of crowd and no one will attend you for qui…	Negative Review	Bad Service

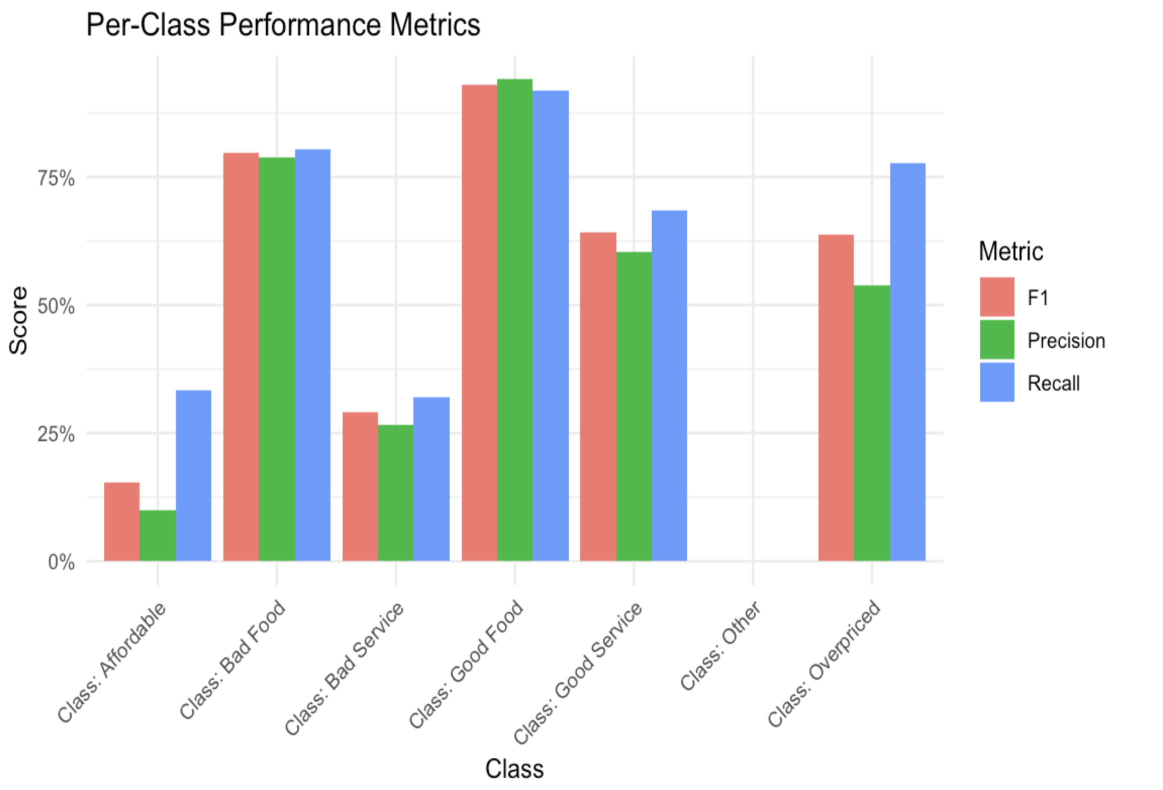
Sentiment Model 2 (Classification)

Reviewer	Restaurant	Rating	Review	Sentiments	Sentiment_Description
Jay Mehta	Labonel	5.0	got a box as a gift! Trully Amazing brownies gotta ex…	Positive Review	Good Food
Vedant Killa	Labonel	5.0	A true match for Theo's from Mumbai, one of Hydera…	Positive Review	Good Food
Anusha Sinha	Driven Cafe	4.0	The cafe is inside Kapil towers in financial district. A…	Positive Review	Good Food
Gourmet Hunter	Driven Cafe	4.0	Good cafe to chill amidst the chaos at this junction. E…	Positive Review	Good Food
Vedant Killa	Driven Cafe	4.0	The ambience is quite a steal! Such interiors and thos…	Positive Review	Good Food
Siva Kumar	Driven Cafe	4.0	Well been to this place for 4 times now, the reason be…	Positive Review	Good Service
Aman Agarwal	Faasos	5.0	Wraps, pancakes, sides, order whatever from here an…	Positive Review	Good Food
Namit Agarwal	Faasos	5.0	I just love the wraps of Faasos and can have them at a…	Positive Review	Good Food

Key Visualization



Sentiment Model 1 (Positive/Negative)



Sentiment Model 2 (Classification)

Key Findings

- **Key Findings from EDA:**
- Evening hours had the highest review frequency.
- **Sentiment Model Insights:**
- **Accuracy:** 95.09%, **Kappa:** 0.81
- **Sensitivity (Positive Reviews):** 97.85%
- **Aspect-Based Classification:**
- Reviews categorized into:
  - Good/Bad Food
  - Good/Bad Service
  - Affordable/Overpriced
- Adds interpretability beyond basic sentiment.

Conclusion

Strengths & Limitations:

•Accurate, interpretable, and ready for real-time.
•Misses sarcasm or complex sentence structures

Real-World Impact:

•Helps restaurants point out exact improvement areas.
•Valuable for marketing, quality control, and customer satisfaction.

References:

1. SquareMeal (2024) *London restaurants that have closed in 2024*. [online] Available at: https://www.squaremeal.co.uk/restaurants/news/london-restaurants-closed-2024_10675 [Accessed 03 Apr. 2025].
2. Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5). Available at: https://www.jstatsoft.org/article/view/v028i05 [Accessed 07 Apr. 2025].4.
3. Salih, A.A. and Abdalla, A.N. (2023) *YOLO model determination flowchart*. Available at: https://www.researchgate.net/figure/YOLO-model-determination-flowchart_fig2_378346802 (Accessed: 10 April 2025).
4. Tidytextmining.com, n.d. *Tidytext Documentation*. [online] Available at: https://www.tidytextmining.com/ [Accessed 12 Apr. 2025].
5. UCI Machine Learning Repository, n.d. *Sentiment Labelled Sentences Data Set*. [online] Available at: https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences [Accessed 16 Apr. 2025].
6. **Feinerer, I.,** 2023. *tm: Text mining package*. [R package] CRAN. Available at: https://cran.r-project.org/package=tm [Accessed 13 Apr. 2025].

Dataset Link (10,000 Restaurant Reviews): https://www.kaggle.com/datasets/joebeachcapital/restaurant-reviews