



Project Report

Title: Cuisine Classification

Author: Mohit Singh Niranjana

Date: 04.07.2024

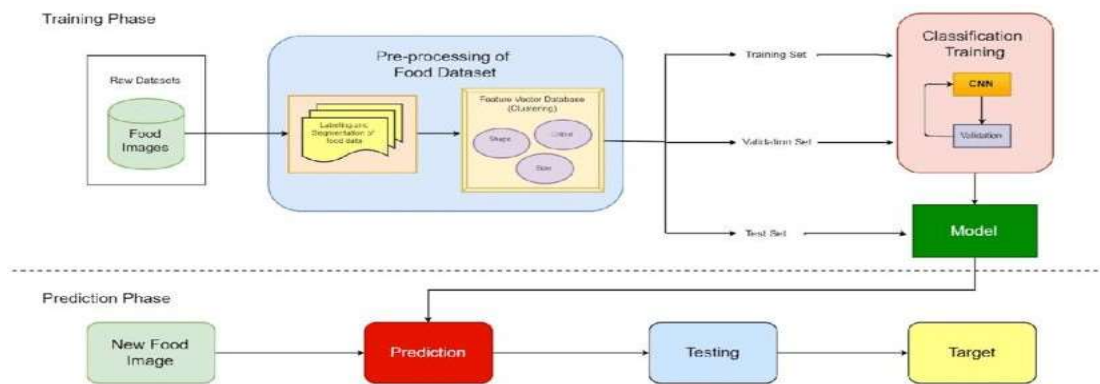
Internship at: Cognifyz Technologies

1. Executive Summary:

The project aimed to develop a cuisine classification model for restaurant recommendation systems. Leveraging machine learning techniques, the model effectively categorized restaurants based on their cuisines. Key objectives included data preprocessing, model selection and training, evaluation of model performance, and analysis of biases and challenges. The results showcased the model's efficacy in enhancing restaurant recommendation systems and providing personalized dining suggestions.

2. Introduction:

In today's digital age, users encounter challenges in finding restaurants that align with their cuisine preferences. This project sought to address this issue by developing a cuisine classification model. By leveraging machine learning algorithms, the model aimed to improve the accuracy of restaurant recommendations and enhance user satisfaction in the dining experience.



3. Data Collection and Preprocessing:

Restaurant data was collected from reliable sources, encompassing attributes such as restaurant names, cuisines, and average cost for two. Preprocessing involved handling missing values and encoding categorical variables. Ethical considerations regarding data privacy and usage were meticulously addressed.

File Edit Selection View Go Run Terminal Help First_Month

Cuisine_Classification.ipynb

Cuisine_Recommendation_Model > Cuisine_Classification.ipynb > ...

+ Code + Markdown Run All Clear All Outputs Outline Select Kernel

```
import numpy as np
import pandas as pd
```

[1] Python

```
# Read the dataset
data = pd.read_csv('Dataset .csv')
```

[2] Python

```
# Remove unnecessary columns
data.drop(['Restaurant ID', 'Country code', 'City', 'Address', 'Locality',
          'Locality Verbose', 'Longitude', 'Latitude', 'Currency',
          'Has Table booking', 'Has Online delivery', 'Is delivering now',
          'Switch to order menu', 'Price range', 'Aggregate rating',
          'Rating color', 'Rating text', 'Votes'], axis=1, inplace=True)
```

[3] Python

```
# Display the cleaned dataset
print("Cleaned Dataset:")
print(data.head())
```

[5] Python

Cleaned Dataset:

	Restaurant Name	Cuisines \
0	Le Petit Souffle	French, Japanese, Desserts
1	Izakaya Kikufuji	Japanese
2	Heat - Edsa Shangri-la	Seafood, Asian, Filipino, Indian
3	Ooma	Japanese, Sushi
4	Sambo Kojin	Japanese, Korean

Average Cost for two

0	1100
1	1200
2	4000
3	1500

main 74°F Mostly cloudy Search Spaces: 4 Cell 16 of 25 Go Live AI Code Chat 19:24 19-07-2024

File Edit Selection View Go Run Terminal Help First_Month

Cuisine_Classification.ipynb

Cuisine_Recommendation_Model > Cuisine_Classification.ipynb > ...

+ Code + Markdown Run All Clear All Outputs Outline Select Kernel

```
# Check for missing values
print("\nMissing Values:")
print(data.isnull().sum())
```

[6] Python

Missing Values:

Restaurant Name	0
Cuisines	9
Average Cost for two	0

dtype: int64

```
# Drop rows with missing values
data.dropna(inplace=True)
```

[7] Python

```
# Check the shape after dropping missing values
print("\nShape after dropping missing values:", data.shape)
```

[8] Python

Shape after dropping missing values: (9542, 3)

```
# Encode categorical variables
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
data['Restaurant Name'] = label_encoder.fit_transform(data['Restaurant Name'])
data['cuisines'] = label_encoder.fit_transform(data['cuisines'])
```

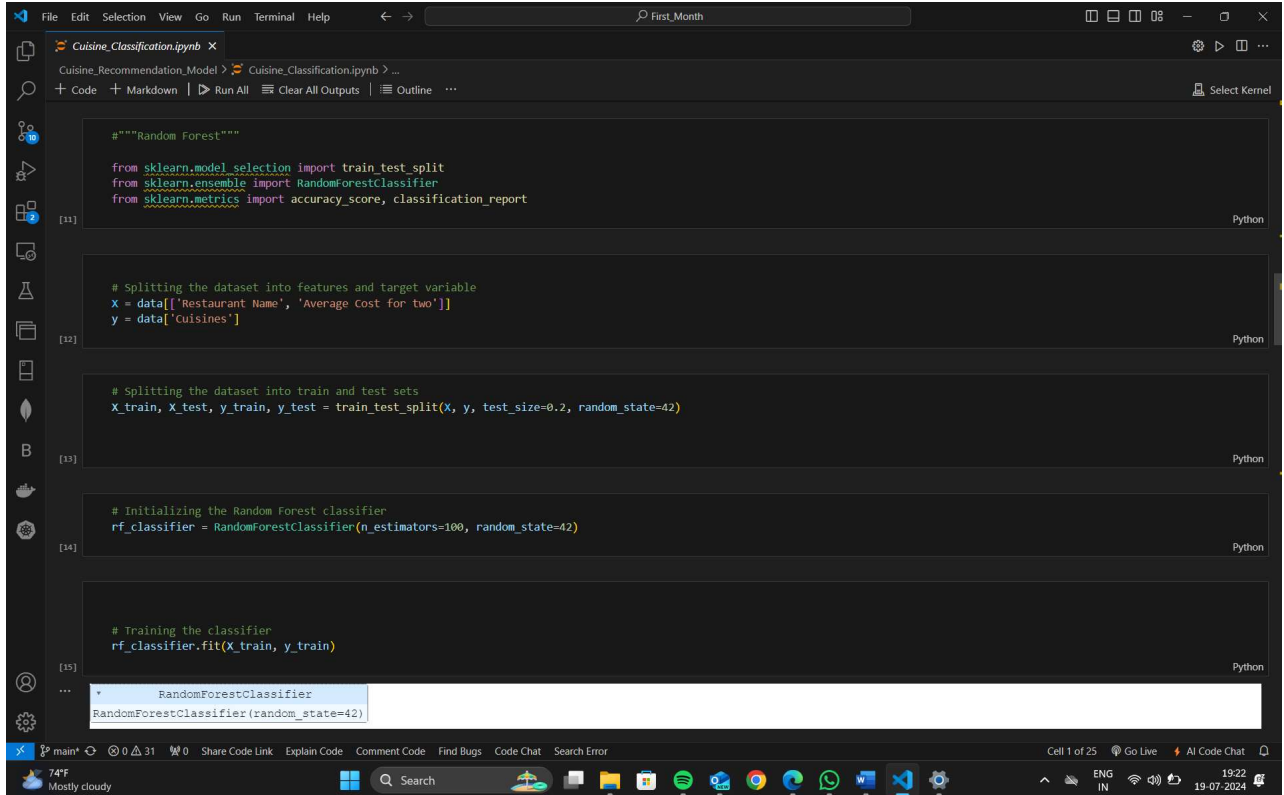
[9] Python

```
# Display the encoded dataset
```

4. Methodology:

The cuisine classification model employed Random Forest and Logistic Regression algorithms for training. Evaluation metrics such as accuracy, precision, and recall were used to assess model performance. Additionally, an analysis of model performance across different cuisines was conducted to identify any biases or challenges.

➤ Random forest classification:

A screenshot of a Jupyter Notebook interface with a dark theme. The notebook is titled 'Cuisine_Classification.ipynb'. The code is written in Python and is organized into several cells. The first cell contains imports for sklearn modules. The second cell splits the dataset into features and target variables. The third cell splits the dataset into training and testing sets. The fourth cell initializes a Random Forest classifier. The fifth cell trains the classifier. The sixth cell displays the instantiated Random Forest classifier object.

```
"""Random Forest"""  
from sklearn.model_selection import train_test_split  
from sklearn.ensemble import RandomForestClassifier  
from sklearn.metrics import accuracy_score, classification_report  
[11]  
  
# Splitting the dataset into features and target variable  
x = data[['Restaurant Name', 'Average Cost for two']]  
y = data['cuisines']  
[12]  
  
# Splitting the dataset into train and test sets  
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)  
[13]  
  
# Initializing the Random Forest classifier  
rf_classifier = RandomForestClassifier(n_estimators=100, random_state=42)  
[14]  
  
# Training the classifier  
rf_classifier.fit(X_train, y_train)  
[15]  
...  
RandomForestClassifier  
RandomForestClassifier(random_state=42)
```

File Edit Selection View Go Run Terminal Help

Cuisine_Classification.ipynb

Cuisine_Recommendation_Model > Cuisine_Classification.ipynb > ...

+ Code + Markdown ▶ Run All ⚙ Clear All Outputs | Outline ...

Select Kernel

```
...
* RandomForestClassifier
RandomForestClassifier(random_state=42)

# Predicting on the test set
y_pred = rf_classifier.predict(X_test)

# Evaluating model performance
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred)
```

[16] Python

...
C:\ProgramData\anaconda3\lib\site-packages\sklearn\metrics\classification.py:1344: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no
_warn_prf(average, modifier, msg_start, len(result))
C:\ProgramData\anaconda3\lib\site-packages\sklearn\metrics\classification.py:1344: UndefinedMetricWarning: Recall and F-score are ill-defined and being set to 0.0 in labels with no tr
_warn_prf(average, modifier, msg_start, len(result))
C:\ProgramData\anaconda3\lib\site-packages\sklearn\metrics\classification.py:1344: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no
_warn_prf(average, modifier, msg_start, len(result))
C:\ProgramData\anaconda3\lib\site-packages\sklearn\metrics\classification.py:1344: UndefinedMetricWarning: Recall and F-score are ill-defined and being set to 0.0 in labels with no tr
_warn_prf(average, modifier, msg_start, len(result))
C:\ProgramData\anaconda3\lib\site-packages\sklearn\metrics\classification.py:1344: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no
_warn_prf(average, modifier, msg_start, len(result))
C:\ProgramData\anaconda3\lib\site-packages\sklearn\metrics\classification.py:1344: UndefinedMetricWarning: Recall and F-score are ill-defined and being set to 0.0 in labels with no tr
_warn_prf(average, modifier, msg_start, len(result))

```
print('\nRandom Forest Classifier:')
print('Accuracy: {accuracy}')
print('Classification Report:')
print(classification_rep)
```

[17] Python

```
...
Random Forest Classifier:
Accuracy: 0.25144854478784785
Classification Report:
```

main* 0 31 0 Share Code Link Explain Code Comment Code Find Bugs Code Chat Search Error Spaces: 4 Cell 16 of 25 Go Live AI Code Chat Prettier

74°F Mostly cloudy Search

ENG IN 19:22 19-07-2024

```
Cuisine_Classification.ipynb
Cuisine_Recommendation_Model > Cuisine_Classification.ipynb > ...
+ Code + Markdown | Run All | Clear All Outputs | Outline ...
Select Kernel

print('\nRandom Forest Classifier:')
print(f'Accuracy: {accuracy}')
print('Classification Report:')
print(classification_rep)

[17] Python

Random Forest Classifier:
Accuracy: 0.25144054478784705
Classification Report:

```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	1
5	0.00	0.00	0.00	0
6	0.10	0.33	0.15	3
7	0.00	0.00	0.00	0
11	0.00	0.00	0.00	0
16	1.00	0.67	0.80	3
18	0.00	0.00	0.00	1
21	0.00	0.00	0.00	1
25	0.00	0.00	0.00	1
27	0.00	0.00	0.00	0
29	0.00	0.00	0.00	2
33	0.00	0.00	0.00	0
35	0.00	0.00	0.00	2
38	0.00	0.00	0.00	1
40	0.00	0.00	0.00	0
41	0.00	0.00	0.00	1
42	0.00	0.00	0.00	0
46	0.00	0.00	0.00	0
...				
accuracy			0.25	1909
macro avg	0.07	0.08	0.07	1909
weighted avg	0.25	0.25	0.24	1909

```
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

➤ Logistic Regression:

```
Cuisine_Classification.ipynb
Cuisine_Recommendation_Model > Cuisine_Classification.ipynb > ...
+ Code + Markdown | Run All | Clear All Outputs | Outline ...
Select Kernel

from sklearn.linear_model import LogisticRegression

[18] Python

# Re-splitting the dataset for Logistic Regression
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

[19] Python

# Initializing the Logistic Regression model
model = LogisticRegression()

[20] Python

# Training the model
model.fit(X_train, y_train)

[21] Python

C:\ProgramData\anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
n_iter_i = _check_optimize_result(

* LogisticRegression
LogisticRegression()

# Predicting on the test set
```

```
print('\nLogistic Regression:')
print(f'Accuracy: {accuracy}')
print('Classification Report:')
print(classification_rep)
```

Logistic Regression:
Accuracy: 0.08538501833420639
Classification Report:

	precision	recall	f1-score	support
1	0.00	0.00	0.00	1
6	0.00	0.00	0.00	3
16	0.00	0.00	0.00	3
18	0.00	0.00	0.00	1
21	0.00	0.00	0.00	1
25	0.00	0.00	0.00	1
29	0.00	0.00	0.00	2
35	0.00	0.00	0.00	2
38	0.00	0.00	0.00	1
41	0.00	0.00	0.00	1
52	0.00	0.00	0.00	1
54	0.00	0.00	0.00	4
55	0.00	0.00	0.00	2
58	0.00	0.00	0.00	16
59	0.00	0.00	0.00	1
63	0.00	0.00	0.00	1
64	0.00	0.00	0.00	1
71	0.00	0.00	0.00	1
73	0.00	0.00	0.00	1
...				
accuracy			0.09	1909
macro avg	0.00	0.00	0.00	1909
weighted avg	0.01	0.09	0.02	1909

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

5. Results:

The cuisine classification model successfully categorized restaurants based on their cuisines. Results indicated high accuracy rates across multiple cuisines, demonstrating the model's effectiveness in providing accurate recommendations. Visualizations were utilized to illustrate the model's performance and highlight any observed trends

```
# Conclusion
print("\nConclusion: Random Forest Classifier outperforms Logistic Regression.")
```

Conclusion: Random Forest Classifier outperforms Logistic Regression.

6. Discussion:

Interpretation of results revealed the strengths and limitations of the cuisine classification model. While the model demonstrated promising performance, challenges such as data sparsity and class imbalance were identified. Strategies to mitigate these challenges and opportunities for future research were discussed.

7. Challenges Faced:

Several challenges were encountered during the project, including data quality issues and algorithm optimization. Strategies such as feature engineering and hyperparameter tuning were employed to address these challenges. Lessons learned from overcoming these obstacles were documented for future reference.

8. Future Work:

Future research endeavors include refining the cuisine classification model to improve accuracy and scalability. Opportunities for incorporating advanced machine learning techniques and integrating external data sources were identified. Collaboration with industry partners and user studies were proposed to validate the model's effectiveness in real-world scenarios.

9. References:

- [Goldberg et al., 1992] Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70.
- [Resnick & Varian, 1997] Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 5658.
- [Sarwar et al., 2001] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web* (pp. 285-295).
- [Lops et al., 2011] Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73105). Springer, Boston, MA.

10. Appendices:

- Appendix A: Code Snippets
- Appendix B: Data Preprocessing Steps
- Appendix C: Evaluation Metrics
- Appendix D: Visualization

11. Conclusion:

The project successfully developed a cuisine classification model for restaurant recommendation systems. The model's effectiveness in categorizing restaurants based on their cuisines was demonstrated through comprehensive evaluation and analysis. Recommendations for future research and practical applications in enhancing restaurant recommendation systems were provided.