

# BREWCAST

## FORECASTING SUCCESS IN THE COFFEE BUSINESS

### MOTIVATION

We believe that predicting successful days will allow a business owner to anticipate when is the day she will need the most staff, best dates to release offers, and overload her stock

### DATASET EXPLANATION

600+ days

Features:

- date
- average temperature
- weekend
- humidity
- rain
- number of customers
- coffee ordered

### DATA ENGINEERING

- Data Collection:
  - Coffee Shop Data: Collected daily sales & customers.
  - Weather Data: Obtained updates from the NMC & Weather Underground
- Data Cleaning:
  - Outlier Detection: removed holidays, ramadan month, and closing days.
- 3. Data Compatibility:
  - Aligned timestamps for accurate analysis.
- 4. Scaling and Transformation:
  - Transformed categorical variables into numerical format for model compatibility.

Avg_Temp	Is_Weekend	Humid	Rainy	Num_Cust	Hot_Drink	Cold_Drink
20.72	0	1	0	88	56	31
21.83	0	0	1	210	116	90
19.02	0	0	0	98	53	44
21.4	0	0	0	125	106	14
16.39	0	0	0	68	54	12
	1	0	0	194	136	56
	1	0	1	167	130	34
	0	1	1	110	74	34
	0	1	0	78	38	37
2	0	1	0	127	89	29
.4	0	1	1	191	84	106
.55	0	1	1	207	128	78
.59	1	0	0	200	162	36
.58	1	0	0	171	110	60
3.41	0	0	0	126	103	18
14.69	0	1	1	130	86	40
21.75	0	0	0	172	91	78

### MODELING WITH RANDOM FOREST

1. Model Selection: Random Forest model for its ability to handle complex relationships and provide robust predictions.
2. Training the Model:
  - Dataset Split: Divided the dataset into training and testing sets to train and evaluate the model's performance.
  - Feature Importance: Evaluated feature importance to understand the contribution of each feature to the predictive power of the model.
  - Data Cleaning: Removed unnecessary redundant data that will not affect the results.
3. Validation and Evaluation:
  - Validation Accuracy: Achieved a robust validation accuracy of 98%, ensuring the model's reliability.
  - Performance Metrics: Utilized mean accuracy rate to comprehensively evaluate the model.
4. Interpretability:
  - Visualizing Trees: Explored individual decision trees within the random forest for interpretability.
5. Practical Insights:
  - Business Impact: Translated model predictions into actionable insights for Maryam, aiding in strategic decision-making.
  - Resource Optimization: provider of information to optimize staffing, inventory, and resource allocation based on anticipated customer demand.

Mean Accuracy: 0.9849228611500701  
Min Accuracy: 0.978494623655914  
Max Accuracy: 0.989247311827957

### RESULTS AND DISCUSSION

The random forest model achieved a 98% validation accuracy, in its efficacy in predicting successful business days for the coffee shop. Crucial features, including temporal patterns and weather conditions, were identified as key contributors. This model provides actionable insights, guiding strategic decisions in promotions and resource allocation. Visualizations enhance interpretability, and despite acknowledging limitations, the real-world application empowers Maryam to optimize operations and enhance overall business success.

## LINK TO DROPBOX

<https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.dropbox.com%2Fsc%2Ffo%2F9nqj3ma7o6wgu6tviqhyc%2Fh%3Frlkey%3Dv8dfdhdhmvld9ks6em09q1mxgf0%26dl%3D0&data=05%7C01%7C202004091%40uaeu.ac.ae%7C545e2d286631427fafc808dbea5a83e2%7C97a92b044c8743419b08d8051ef8dce2%7C0%7C0%7C638361446724557811%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6IjEhaWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=Ojnl%2FreDpGq5G6w9yAjiymwSEllnldwCd4WPnXSF0fo%3D&reserved=0>