CalmSense

1) Load Data

CSV loaded from: /Users/urmebose/Documents/CalmSense/Stress-Lysis.csv

Preview:

	Humidity	Temperature	Step_count	Stress_Level
0	21.33	90.33	123	1
1	21.41	90.41	93	1
2	27.12	96.12	196	2
3	27.64	96.64	177	2
4	10.87	79.87	87	0

Shape: `(2001, 4)`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2001 entries, 0 to 2000
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Humidity      2001 non-null   float64
 1   Temperature   2001 non-null   float64
 2   Step_count    2001 non-null   int64
 3   Stress_Level  2001 non-null   int64
dtypes: float64(2), int64(2)
memory usage: 62.7 KB
```

2) Clean & Rename

Renamed columns:


```
[
  0 : "humidity"
  1 : "temperature"
  2 : "steps"
  3 : "stress_label"
]
```


Dropped 0 rows => shape=(2001, 4)

Noise

☑ Noise (temp/humidity)?

Noise Level (%)

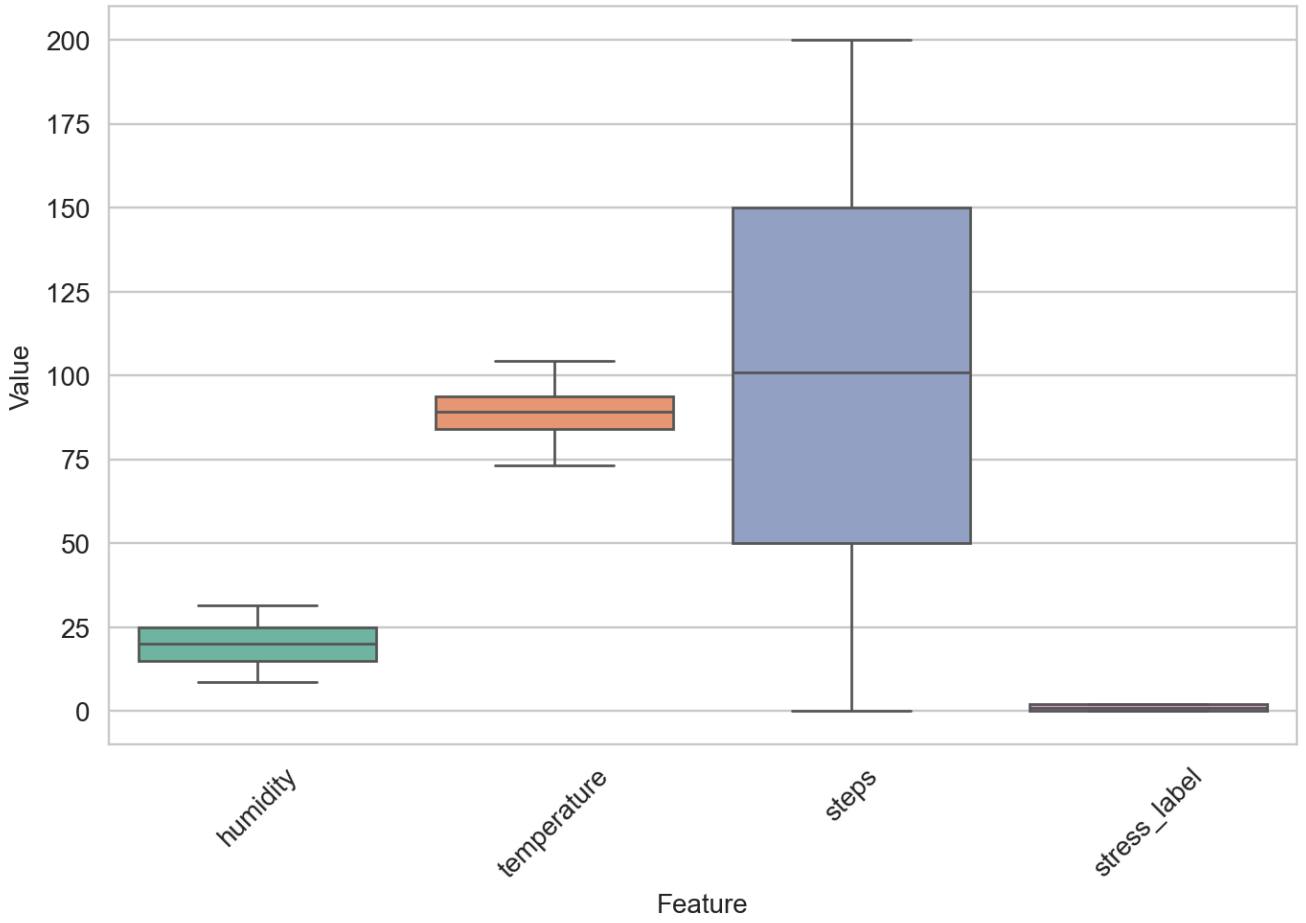
3

1 10

Injected ~3% noise.

3) Advanced Stats & Visuals

Box Plot

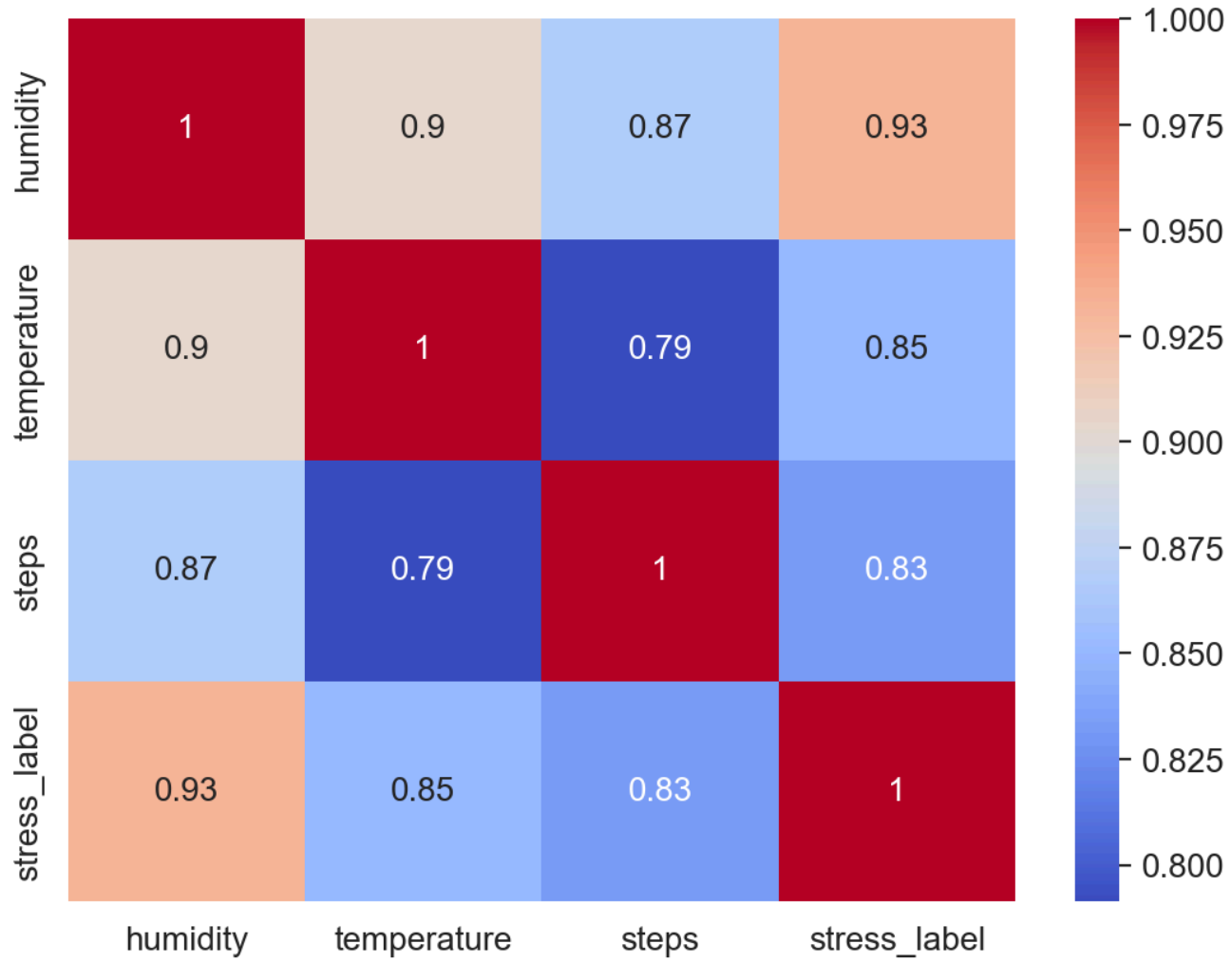


T-test (temp vs humidity): t=362.774, p=0.000

ANOVA => F=2634.345, p=0.000

Correlation Matrix

	humidity	temperature	steps	stress_label
humidity	1	0.9037	0.8665	0.9317
temperature	0.9037	1	0.7913	0.8504
steps	0.8665	0.7913	1	0.8326
stress_label	0.9317	0.8504	0.8326	1



4) Outlier Removal

Removed 40 outliers => shape=(1961, 4)

5) Feature Engineering

Created 'hum_temp_interact' = humidity * temperature

Shape => `(1961, 5)`

6) Split & Scale

Train size=1568, Test size=393

7) Model Tuning (RandomForest + XGBoost)

RandomForest Search

RF best params:


```
{
  "n_estimators" : 50
  "min_samples_split" : 2
  "min_samples_leaf" : 1
  "max_depth" : 5
}
```


RF CV accuracy: `0.9776757043190747`

XGBoost Search

XGB best params:


```
{
  "n_estimators" : 50
  "max_depth" : 3
  "learning_rate" : 0.1
}
```


XGB CV accuracy: `0.9776757043190747`

8) Evaluate & Pick Best

Test-Set Metrics

	Model	Accuracy	Precision	Recall	F1
0	RandomForest	98.22%	0.98	0.98	0.98
1	XGBoost	98.22%	0.98	0.98	0.98

RandomForest => Confusion Matrix & Classification Report

0	1	2
93	3	0
4	154	0
0	0	139

```
precision    recall  f1-score   support

           0       0.96      0.97      0.96        96
           1       0.98      0.97      0.98       158
           2       1.00      1.00      1.00       139

    accuracy                           0.98       393
   macro avg       0.98      0.98      0.98       393
weighted avg       0.98      0.98      0.98       393
```

XGBoost => Confusion Matrix & Classification Report

0	1	2
94	2	0
5	153	0
0	0	139

```
precision    recall  f1-score   support

           0       0.95      0.98      0.96        96
           1       0.99      0.97      0.98       158
           2       1.00      1.00      1.00       139

    accuracy                           0.98       393
   macro avg       0.98      0.98      0.98       393
weighted avg       0.98      0.98      0.98       393
```

Best model is RandomForest, accuracy=98.22%

Final Best Model Accuracy

98.22%

9) Real-Time Simulation

	humidity	temperature	steps	hum_temp_interact	Row #	Prediction
0	14.7519	80.9771	9	1,194.567	1	Low Stress
1	27.8725	96.2871	137	2,683.7608	2	High Stress
2	11.2495	79.3412	23	892.5528	3	Low Stress
3	24.3215	96.2829	164	2,341.7461	4	High Stress
4	23.7866	93.6859	178	2,228.4672	5	High Stress
5	21.4747	89.8427	96	1,929.3403	6	Normal Stress
6	10.7273	77.8231	45	834.8304	7	Low Stress
7	27.3029	98.4067	135	2,686.784	8	High Stress
8	12.2583	82.8776	64	1,015.9407	9	Low Stress
9	17.1182	89.8895	50	1,538.7478	10	Normal Stress

10) Permutation Feature Importance