ORAL TEMPERATURE PREDICTION WITH MACHINE LEARNING

Sara Sarafimova

THE PROCESS

1. Data Exploration
2. Data Preprocessing
3. Data Visualization
4. Model Training
5. Feature Selection
6. Hyperparameter Optimization

Data Exploration

4080 infrared thermograms of subjects' faces

Technology used: FLIR (Forward Looking Infrared)

4 rounds of images taken

1020 subjects

37 features

(26 facial features)

Target: Average Oral Temperature

The Dataset

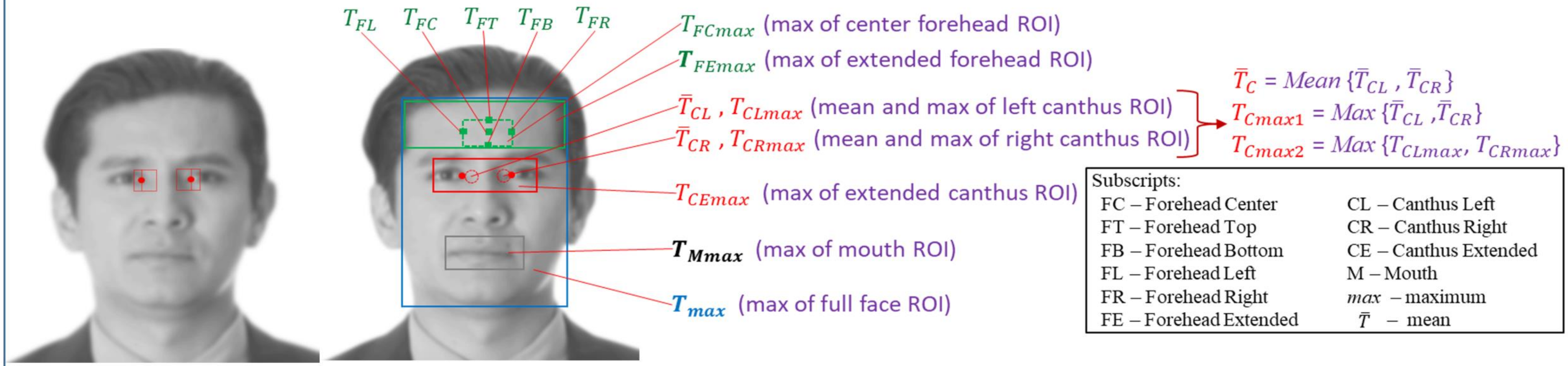


Figure 1. Delineated facial regions and critical points on thermal images: forehead regions and points (green), canthi region and points (red), mouth region (gray rectangle), and entire face (blue rectangle).

Note: The above image is a generic face (based on PowerPoint clip art: Insert > Icons > Cutout People >Alfredo) used for illustration purposes and not an actual participant in our study.

The Dataset

Feature	Region of Interest	Value Calculation
T_RC	24x24 pixels around the right canthus	Average temperature of the highest 4 pixels
T_RC_Dry	16x24 pixels around the right canthus dry area	Average temperature of the highest 4 pixels
T_RC_Wet	8x24 pixels around the right canthus wet area	Average temperature of the highest 4 pixels
T_RC_Max	24x24 pixels around the right canthus	Maximum temperature within the square
T_LC	24x24 pixels around the left canthus	Average temperature of the highest 4 pixels
T_LC_Dry	16x24 pixels around the left canthus dry area	Average temperature of the highest 4 pixels
T_LC_Wet	8x24 pixels around the left canthus wet area	Average temperature of the highest 4 pixels
T_LC_Max	24x24 pixels around the left canthus	Maximum temperature within the square
RCC	3x3 pixels centered at the right canthus point	Average temperature within the square
LLC	3x3 pixels centered at the left canthus point	Average temperature within the square
canti4Max	Extended canthi area	Average temperature of the highest 4 pixels
T_OR_Max	Oral/mouth region	Maximum temperature within the region

The Dataset

Feature	Definition
SubjectID	ID number of the subject
Date	Date of data collection
Round	Round of pictures taken (1, 2, 3, 4)
Age	Age range of the subject
Gender	Male or female
Ethnicity	Ethnicity of the subject
Distance	Distance between the subject and the IRT
Cosmetics	1 – cosmetics applied, 0 – no cosmetics applied
T_atm	Ambient temperature
Humidity	Relative humidity
T_offset	Temperature difference between the set and measured blackbody temperature
aveOralF	Average oral temperature measured twice under fast mode
aveOralM	Average oral temperature measured twice under monitor mode

Data Preprocessing

Handling missing values

- 1917 missing values
- Filling the missing values with the mean of the column for each subject

Feature Encoding

Label Encoder Mappings for Ethnicity

- 0 -> American Indian or Alaskan Native
- 1 -> Asian
- 2 -> Black or African-American
- 3 -> Hispanic/Latino
- 4 -> Multiracial
- 5 -> White

Label Encoder Mappings for Gender

- 0 -> Female
- 1 -> Male

Ordinal Encoding for Age

- "18-20": 0, "21-25": 1, …, "51-60": 6, ">60": 7

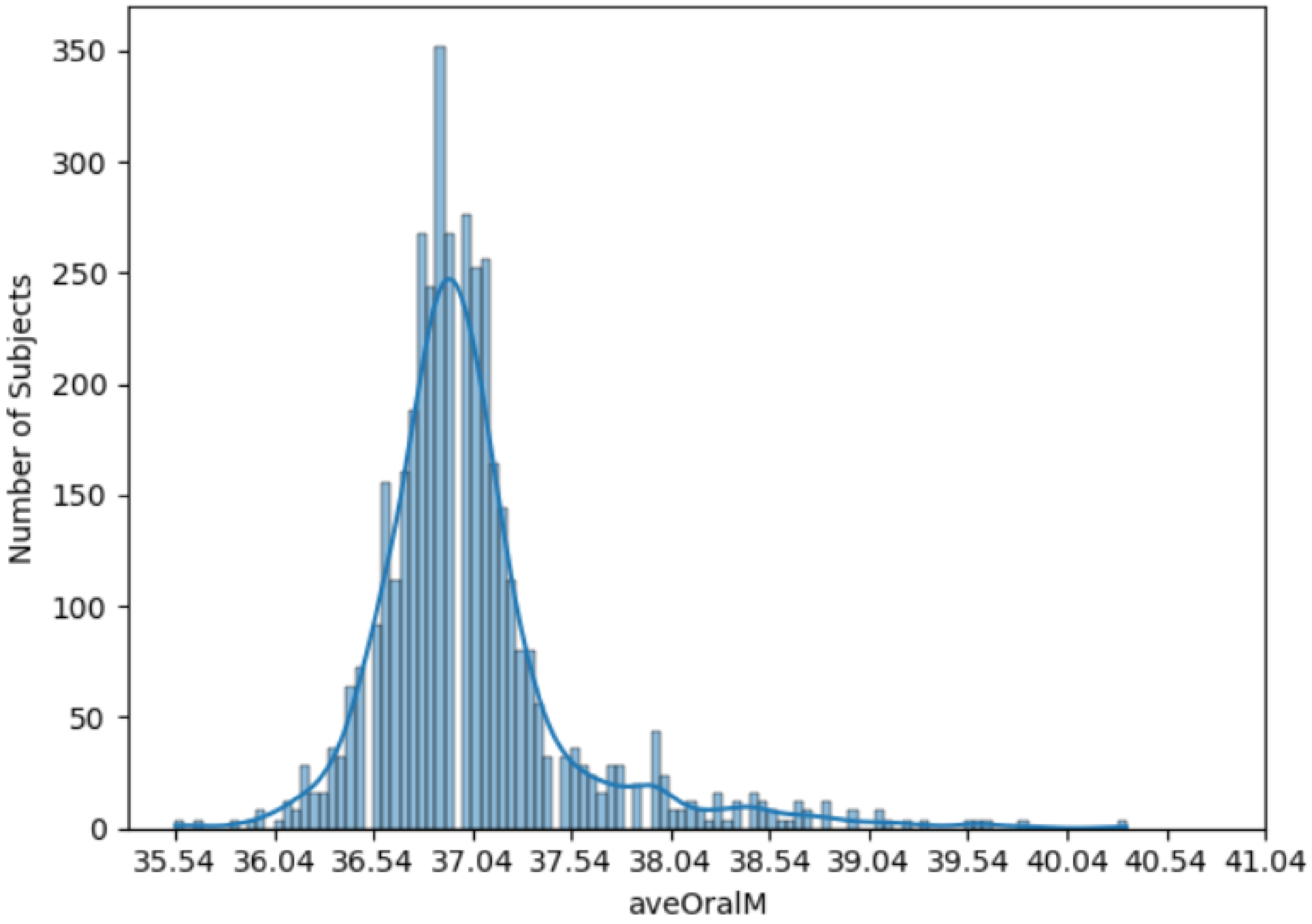
Data Standardization

- Standardizing all* values except the target variables

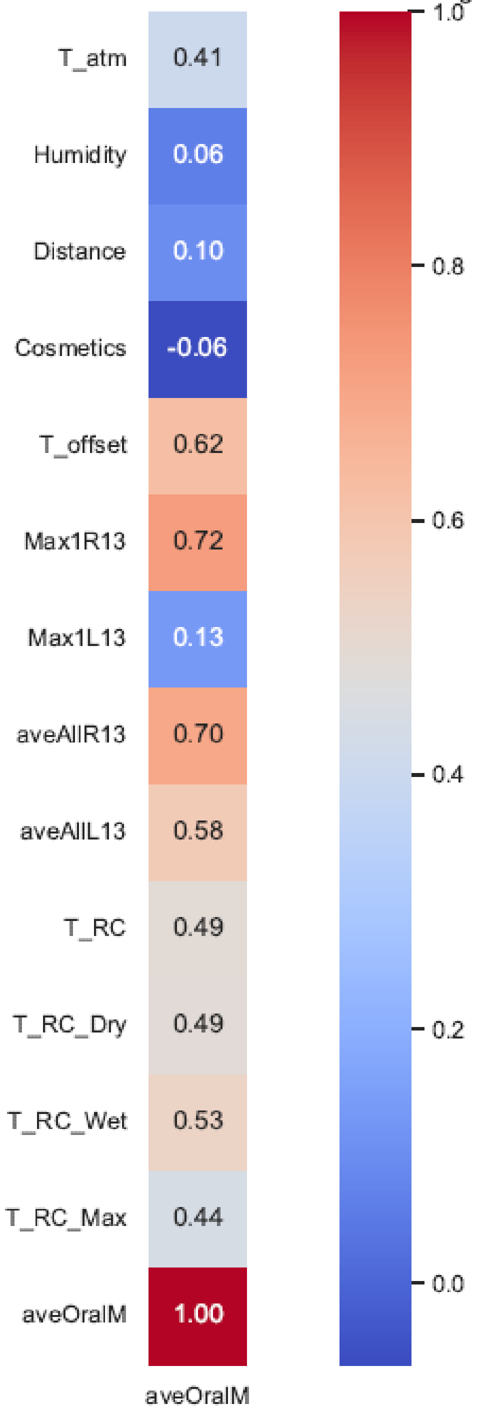
* excluded: SubjectID, Age, Gender, Ethnicity, Cosmetics, Date

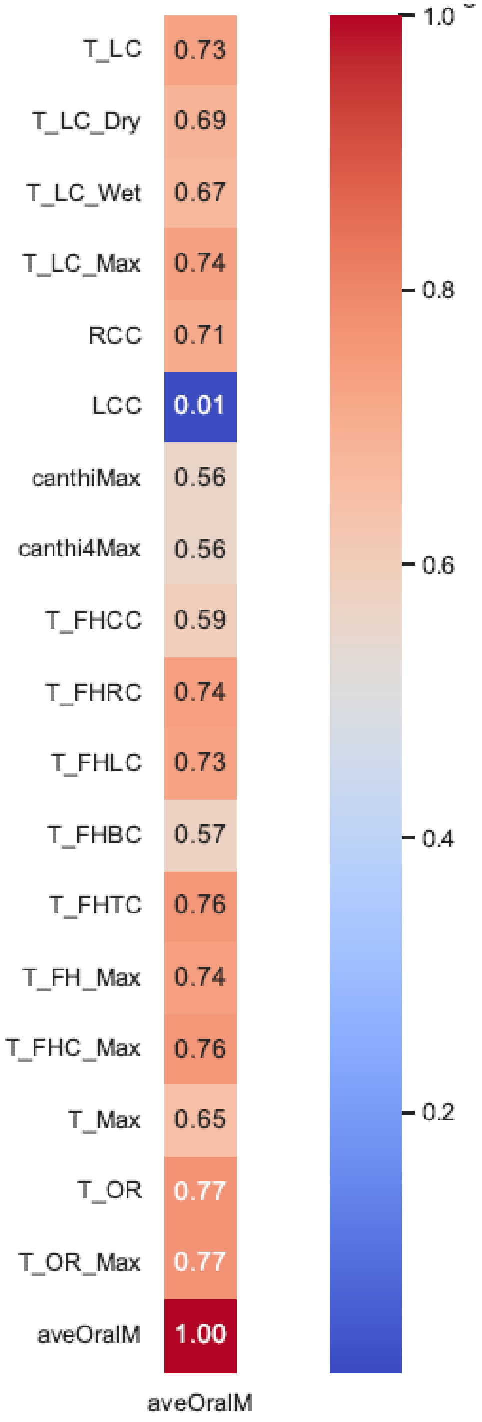
DATA VISUALIZATION

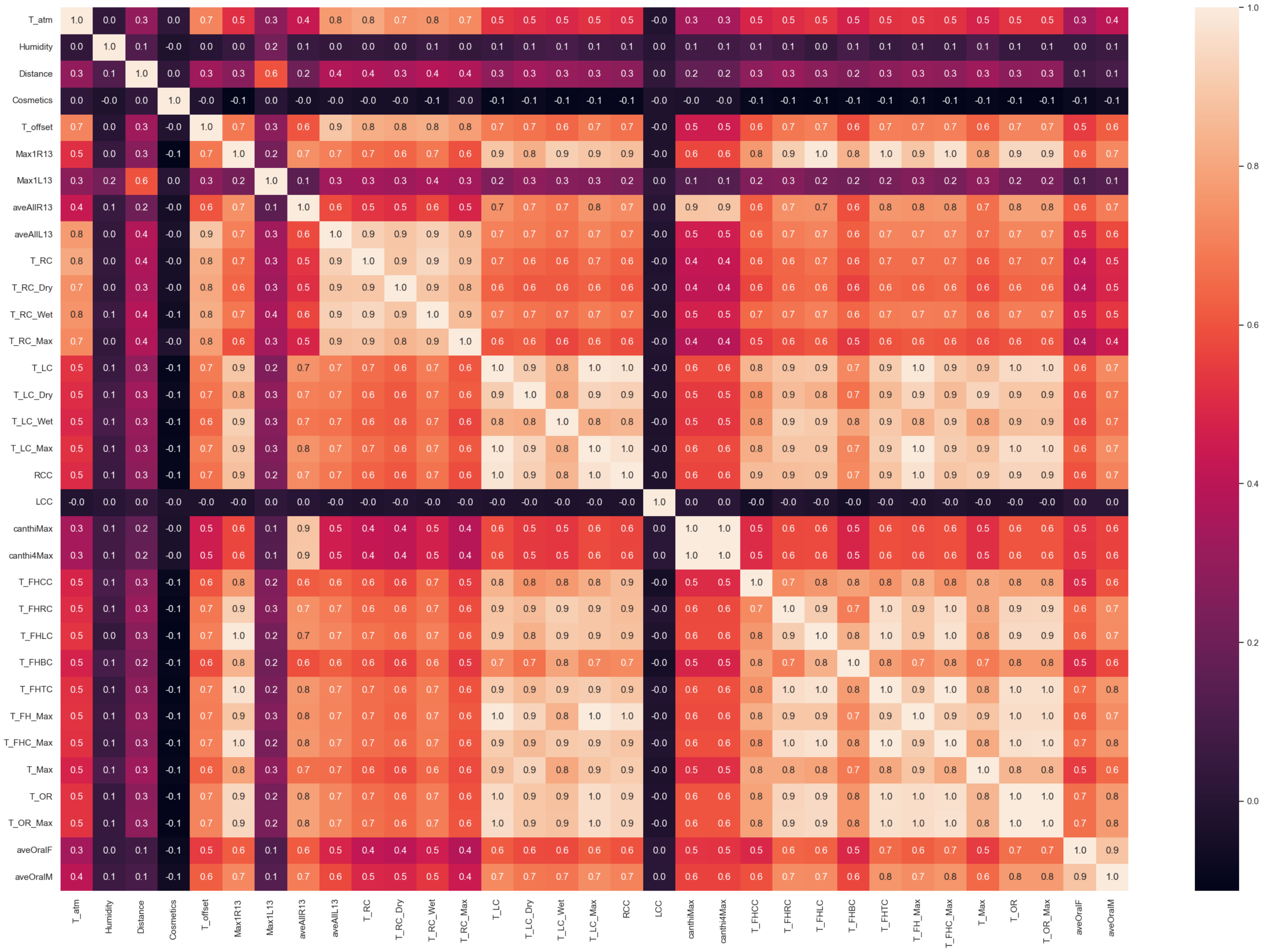
Distribution of the Target Variable



Correlation matrix






T_atm
Humidity
Distance
Cosmetics
T_offset
Max1R13
Max1L13
aveAllR13
aveAllL13
T_RC
T_RC_Dry
T_RC_Wet
T_RC_Max
T_LC
T_LC_Dry
T_LC_Wet
T_LC_Max
RCC
LCC
canthiMax
canthi4Max
T_FHCC
T_FHRC
T_FHLC
T_FHBC
T_FHTC
T_FH_Max
T_FHC_Max
T_Max
T_OR
T_OR_Max
aveOralF
aveOralM
1.0
0.8
0.6
0.4
0.2
0.0

MODEL TRAINING

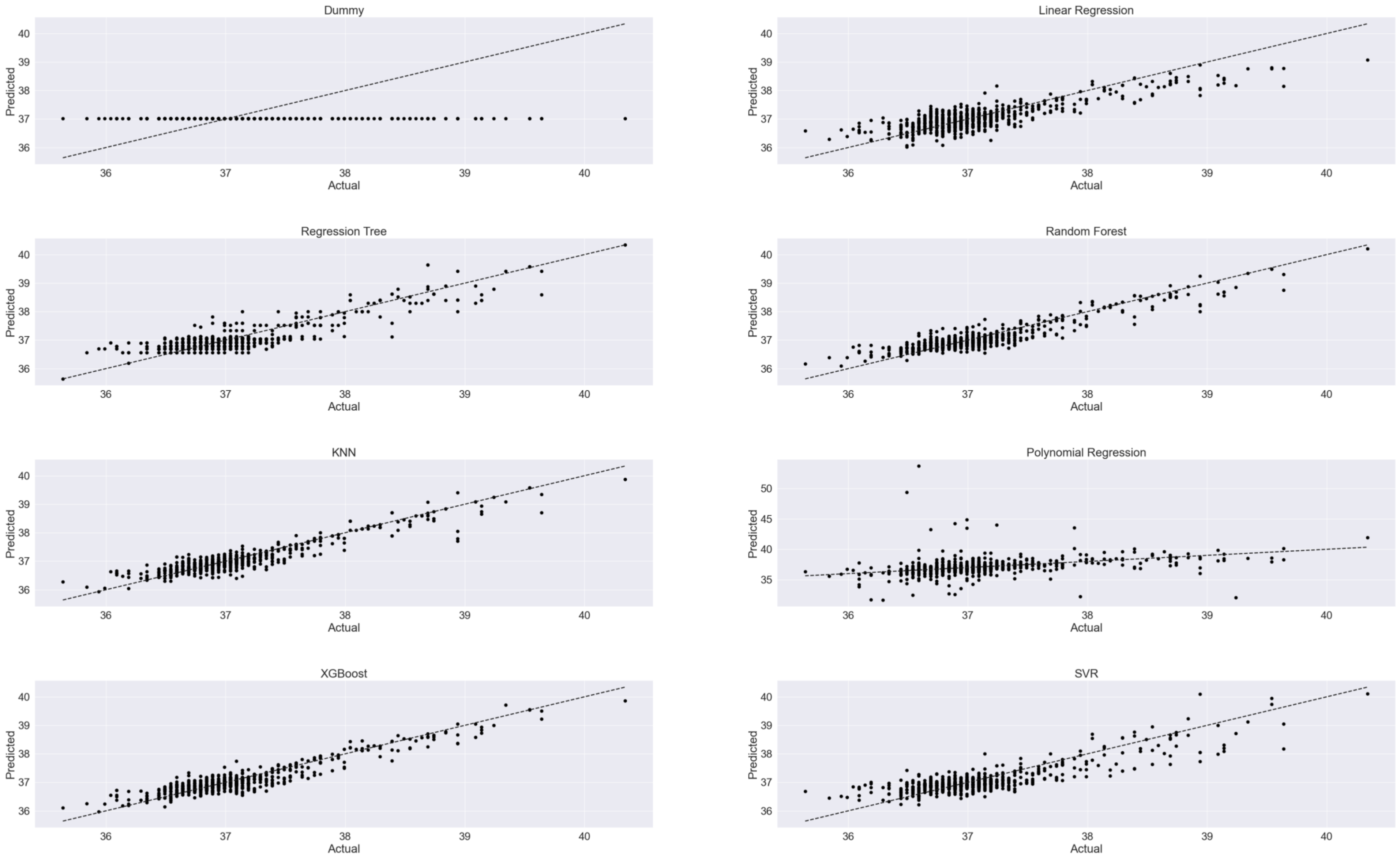
- Dummy Regressor
- Linear Regression
- Regression Tree
- Random Forest Regressor
- K-nearest Neighbors Regressor
- Polynomial Regression with Degree 3
- Extreme Gradient Boost
- SVR

Evaluating the Models

Models with Default Parameters

Metric	Dummy	Dummy 2	Linear Reg	Linear Reg 2	Reg Tree	Reg Tree 2	Random Forest	Random Forest 2	KNN	KNN 2	XGB	XGB 2	SVR	SVR 2
MAE	0.34095	0.33276	0.22503	0.22615	0.20780	0.20703	0.15997	0.15164	0.13619	0.12344	0.15025	0.14863	0.21547	0.21040
MSE	0.29815	0.25934	0.08961	0.08666	0.07680	0.07661	0.04747	0.04240	0.04071	0.03605	0.04052	0.04066	0.08831	0.08633
RMSE	0.54604	0.50827	0.29935	0.29420	0.27713	0.27603	0.21789	0.20533	0.20177	0.18921	0.20130	0.20160	0.29717	0.29363
R2	-0.00702	-0.0008	0.69734	0.65737	0.74062	0.69289	0.83965	0.83258	0.86249	0.85848	0.86314	0.84196	0.70174	0.66557

The comparison is done between the 80-20 data split validation technique (“model name”) and the 15-fold cross-validation validation technique (“model name 2”).



80-20 data split

FEATURE SELECTION

20 best features

Univariate Selection	mRMR
T_offset	T_offset
Max1R13	Max1R13
aveAllR13	aveAllR13
aveAllL13	aveAllL13
T_LC	T_LC
T_LC_Dry	T_LC_Dry
T_LC_Wet	T_LC_Wet
T_LC_Max	T_LC_Max
RCC	RCC
TCEmax	TCEmax
TFC	TFC
TFR	TFR
TFL	TFL
TFB	canthi4Max
TFT	TFT
TFEmax	TFEmax
TFCmax	TFCmax
Tmax	Tmax
T_OR	T_OR
T_OR_Max	T_OR_Max

MODEL TRAINING

on the new dataset with 20 best features

Evaluating the Models

Models with Default Parameters

Metric	Dummy 2	Dummy _20	Linear Reg 2	Linear Reg_20	Reg Tree 2	Reg Tree_20	Random Forest 2	Random Forest _20	KNN 2	KNN_20	XGB 2	XGB_20	SVR 2	SVR_20
MAE	0.33276	0.33291	0.22615	0.24014	0.20703	0.215805	0.15164	0.19472	0.12344	0.20884	0.14863	0.20469	0.21040	0.27936
MSE	0.25936	0.25957	0.08666	0.09586	0.07661	0.083894	0.04240	0.06582	0.03605	0.07811	0.04066	0.07305	0.08633	0.14868
RMSE	0.50827	0.50910	0.29420	0.30953	0.27603	0.289479	0.20533	0.25646	0.18921	0.27930	0.20160	0.27010	0.29363	0.38558
R2	-0.0008	-0.0023	0.65737	0.62844	0.69289	0.674581	0.83258	0.74439	0.85848	0.69646	0.84196	0.71566	0.66557	0.42272

The comparison is done between the 15-fold cross-validation technique on the raw dataset (“model name 2”) and the same validation technique on the dataset with 20-best features (“model name_20”).

Conclusion on Feature Selection

The lack of progress in model performance when selecting only the top 20 features may stem from either overfitting or insufficient information for the model to effectively learn from. Hence, we will use the original dataset with 36 features.

HYPERPARAMETER OPTIMIZATION

Bayesian Optimization

	KNN	XGB	RF
OPTIMIZED PARAMETERS	k = 3 weights = ‘distance’ p = 1 (Manhattan distance)	alpha = 0.4 booster = gbtree eta = 0.141 max_depth = 8 min_child_weight = 2	/

Metric	KNN 2	KNN BO	XGB 2	XGB BO
MAE	0.12344	0.10384	0.14863	0.14164
MSE	0.03605	0.02967	0.04066	0.03725

BO - Bayesian Optimization

Conclusion on Hyperparameter Optimization

Considering the slight improvements in errors, the effort and time taken to tune the parameters, it is debatable whether Bayesian Optimization was worth it. Since the time was not too long and there are improvements after all, we will suppose it paid off.

CONCLUSION

- The 15-fold CV yielded better results than the 80/20 percentage split
- KNN proved to be the best model throughout the whole process, with a MAE of 0.10384 and a MSE of 0.2967 after hyperparameter optimization
- Experiments demonstrate that integrating feature selection and hyperparameter optimization, in this particular problem, sadly did not yield significant improvement.