Journal of Physics:
Conference Series

PAPER • **OPEN ACCESS**

The Cool, Quiet City machine learning competition: Overview and results

To cite this article: C Miller *et al* 2025 *J. Phys.: Conf. Ser.* **3140** 112017

The Cool, Quiet City machine learning competition: Overview and results

C Miller[1,2,*], M Ibrahim[3], I S Akbar[4], B Picchetti[1], Y X Chua[1], M Frei[1,5], F Biljecki[1], A Chong[1], M Quintana[1,6], C Fu[1,7]

[1] College of Design and Engineering, National University of Singapore (NUS), Singapore
[2] College of Integrative Studies, Singapore Management University (SMU), Singapore
[3] Cairo University, Egypt
[4] Ministry of Finance, Indonesia
[5] Berkeley Education Alliance for Research in Singapore (BEARS), Singapore
[6] Future Cities Lab Global, Singapore-ETH Centre (SEC), Singapore
[7] Centrica Business Solutions, Antwerp, Belgium

E-mail: *`cmiller@smu.edu.sg`

Abstract. The prediction of thermal and noise-based preferences in the urban context is valuable in characterizing interventions to mitigate the challenges of health, productivity, and satisfaction of urban dwellers. The growth of crowd-sourced data and data-driven techniques provides an opportunity to increase the understanding of which machine learning models are most accurate and applicable for this context. This paper outlines the results of a machine learning competition aiming to enhance the accuracy of predicting human comfort in the city context. The competition asks contestants to use contextual data to predict noise distraction and thermal preference in various indoor and outdoor spaces. This competition included the city-scale collection of 9,808 smartwatch-driven micro-survey responses that were collected alongside 2,659,764 physiological and environmental measurements from 98 people using an open-source watch-based platform combined with geolocation-driven urban digital twin metrics. This paper explains the two best solutions to this competition and provides a discussion of the factors that may have contributed to their accuracy of more than 0.7 in multiclass tasks. These solutions notably included the use of LightGBM, XGBoost, CatBoost, and simple Neural Networks while avoiding overly complex solutions such as deep learning or recurrent architectures, which offer limited advantages for structured data classifications.

1. Introduction

Urban areas face the growing challenge of noise and heat, which affect our comfort, health, and productivity. Effectively managing these factors is crucial because they influence people's productivity, satisfaction, and stress levels in indoor and outdoor environments. Knowing and predicting locations with comfortable thermal and acoustic conditions empowers people to find refuge from heat and noise. This knowledge could relieve the negative impacts, making the city appear cooler and quieter while addressing personal comfort preferences. People are complex in the aspects of the acoustic and environmental environment that satisfy them. The greatest dissatisfaction with indoor office spaces is noise privacy and thermal comfort [1]. Field-based data collection studies are growing to characterize the sources and mitigation of these aspects of human experience [2]. Large data sets and machine learning competitions have been used in

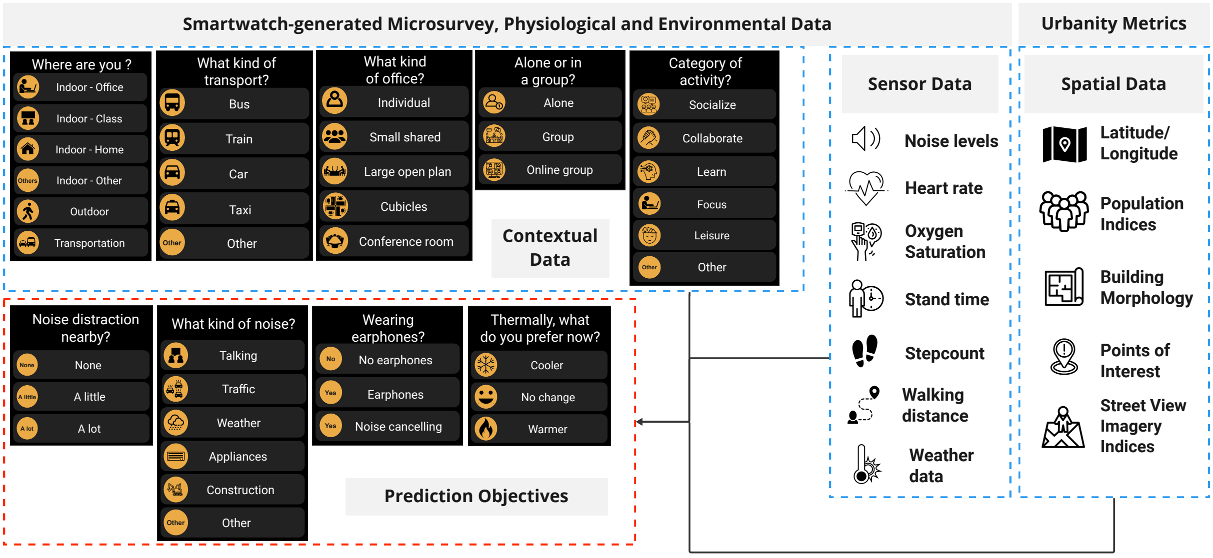

Figure 1. Training/testing data and prediction objectives structure for the Cool, Quiet City Competition (adopted from [8]).

research to address the *generalizability* challenge [3] and the best models for specific applications [4, 5].

This paper outlines a 2023 machine learning competition to test the ability of various data-driven methods to predict noise and heat issues in the tropical context of Singapore. This competition had a total of 23 participants who made up 16 teams and who submitted 222 prediction submissions during the competition, which lasted from November 15, 2023, to February 23, 2024. The competition focused on self-reported large-scale heat and noise data in a city combined with contextual environmental, physiological, and urban-scale digital twin context training data outlined in previous work [6, 7]. The introduction of this competition was described in a previous work [8], of which some content is adapted and replicated in this overview of the results.

2. Methodology

The data collection process for this competition used the open source Cozie Apple platform [7, 9]. This mobile and smartwatch application for iOS devices allows people to complete a watch-based micro-survey and provide real-time feedback about environmental conditions via their Apple Watch. Using the built-in sensors of the smartwatch, it collected physiological (e.g. heart rate, activity), environmental (sound level), and location (latitude and longitude) data. The deployment for this work includes the watch microsurvey responses from 106 people, who each provided at least 100 micro-survey responses from October 2022 to August 2023 [6]. For each survey response location, the spatial indicators created by the digital twins were calculated using the Urbanity Python package [10, 11]. The location (longitude and latitude) was recorded when a micro-survey response was submitted. In addition to the micro-survey, data on heart rate, noise level, blood oxygen saturation, standing time, step count, and walking distance were collected. Air temperature (°C) and relative humidity (%) data were collected from 13 weather stations.

The Cool, Quiet City Competition was hosted as a Kaggle Community Competition at the end of 2023 [12]. The data set described previously was divided into training and public and private leaderboard test data segments [8]. The training data included all variables, while testing and validation separated the prediction objectives into a subset that was not provided to the contestants. Figure 1 outlines the various data types included in the competition training and test data sets. The competition participants used the Kaggle platform to download the training data for model development and then upload the prediction objective variables for the test data sets to receive a classification accuracy metric score instantly computed to show their rank on a public leaderboard. The contestants then iterated their method and reuploaded the predictions throughout the competition, and the private leaderboard was calculated at the conclusion.

3. Results and Discussion

The paper outlines the results of the two top solutions in the competition and outlines the machine learning workflow, modeling characteristics, and lessons learned.

3.1. The First-Place Solution

The highest ranking solution was created by Michael Ibrahim [13] from Cairo University and its architecture is shown in Figure 2. The top-ranked solution employs a robust machine learning pipeline that integrates feature engineering, ensemble modeling, and post-processing. It combines three gradient-boosting algorithms, LightGBM [14], XGBoost [15], and CatBoost [16], using stratified 5-fold cross-validation and soft voting. No hyperparameter tuning was performed, using the default parameters. Additionally, a rule-based postprocessing step is applied to resolve label inconsistencies, ensuring that the predictions adhere to domain-specific logical constraints. The implementation of the top-ranked solution is available in a Kaggle notebook [17].

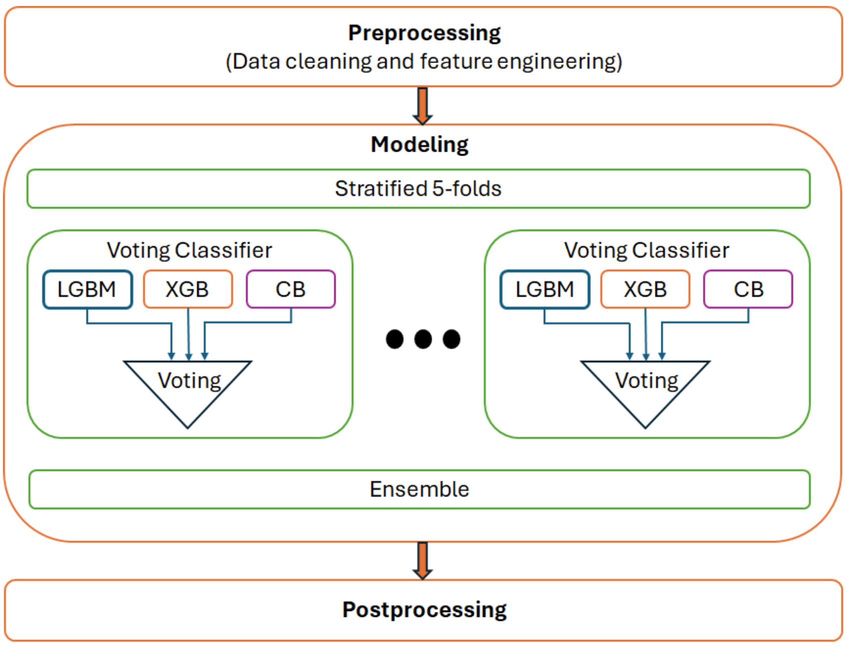

Figure 2. Modeling architecture for the first place solution

The data preprocessing phase of the pipeline incorporated multiple strategies to enhance the performance of the model. Cyclical feature encoding was applied to wind direction, decomposing it into sine and cosine components to capture cyclical patterns, while time-based features such as hour and minute were similarly transformed using trigonometric functions. Categorical variables, such as *q_location*, were processed using a hot encoding, whereas the ordinal encoding converted the survey responses into numerical values. Temporal features were refined by logarithmically transforming the duration of survey completion and deriving time-of-day and day-of-week indicators, which were then encoded cyclically to account for periodicity. Physiological data, including heart rate and step count, were normalized per participant using median, mean, and standard deviation, with linear interpolation ensuring alignment with survey timestamps. Furthermore, weather data from the nearest meteorological station, including temperature, humidity, and wind speed/direction, were mapped to survey responses using the Haversine distance formula and missing values were imputed with temporal proximity-based checks.

Then, for each fold $K \in \{1, \ldots, 5\}$, three gradient-boosted tree models are trained with different algorithmic strategies. The probabilistic outputs of these models are aggregated using a soft voting ensemble, where the final class probability for the fold K is calculated as the arithmetic mean:

$$P_K(y = c) = \frac{1}{3} \sum_{m=1}^{3} P_{m,K}(y = c)$$

where $P_{m,K}$ denotes the probability output of model m in the fold K. This approach balances the strengths of each algorithm, mitigating the variance of the individual model through probabilistic consensus[18].

The fold-level voting classifiers are further combined into a hierarchical ensemble by averaging their probabilistic predictions:

$$P_{final}(y = c) = \frac{1}{5} \sum_{K=1}^{5} P_K(y = c)$$

This simple (i.e., unweighted) averaging aggregation ensures equal contribution from all folds, promoting model robustness and simplicity while mitigating fold-specific biases [19].

The final predictions undergo domain-specific postprocessing to resolve logical inconsistencies identified during the exploratory analysis. For example, respondents reporting *A little* ambient noise (q_noise_nearby) alongside *Not applicable* for noise type (q_noise_kind) and earphone usage (q_earphones) are adjusted to *No earphones*, reflecting the plausible inference that the absence of earphones explains their perception of ambient noise. Mutual exclusivity rules are enforced to prevent contradictory or inconsistent response combinations. Post-processing scripts iteratively apply these corrections to ensure adherence to domain logic. This step bridges statistical predictions with real-world plausibility, improving the practical utility of the model outputs.

3.2. The Second-Place Solution

The second place solution was created by Iqbal Syah Akbar [20] of the Indonesia Ministry of Finance, and its structure is shown in Figure 3. This solution involves multiple data preprocessing steps and a stacked ensemble consisting of logistic regression, support vector machine (SVM), LightGBM, XGBoost, and CatBoost, with a neural network serving as the meta-estimator. With the exception of CatBoost, all were trained using their default settings. The implementation of the second-place solution is available in a Kaggle notebook [21].

The first step is pre-processing and feature engineering, where all survey timestamps are decomposed into year, month, week, weekday, day, hour, minute, second, and millisecond components. Additionally, the time difference between consecutive timestamps is calculated to account for cases where the location timestamp is recorded only after the survey submission. Finally, each participant's location is assigned to the nearest station using the Euclidean distance.

The features are then aggregated starting with the weather station data, which are aggregated into mean and standard deviation on a minute-by-minute basis. Survey counts are aggregated at the daily, hourly, and minute levels. Finally, activity data, such as heart rate and oxygen saturation, are aggregated using mean and standard deviation.

Journal of Physics: Conference Series **3140** (2025) 112017 doi:10.1088/1742-6596/3140/11/112017

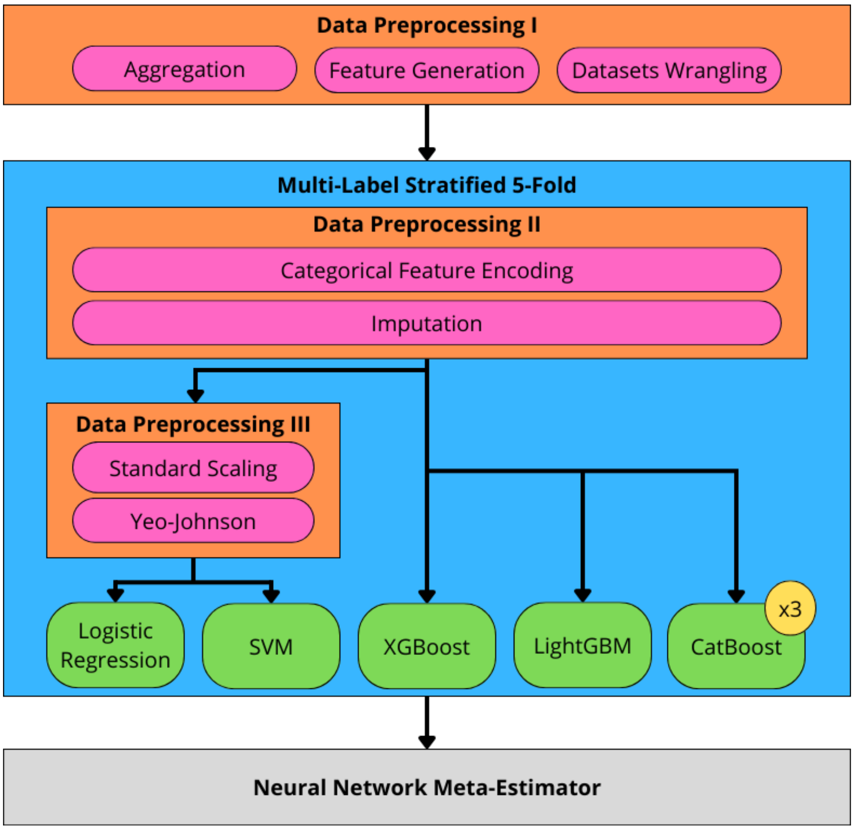

Figure 3. Modeling architecture for the second place solution

For cross-validation, the solution employs a 5-fold stratified multi-label approach, which combines k-fold splitting with iterative stratification to preserve the characteristics of the data; the dataset is divided into k folds, with each fold serving as a validation set once while the remaining folds are used for training. Iterative stratification ensures that the proportion of each class is maintained across all labels during the split process [22].

To ensure high accuracy, the solution uses an ensemble with diverse models. Logistic regression and support vector machines (SVM) are models that, despite their low accuracy, provide high diversity to the ensemble. These learners use mean imputation, m-estimate encoding [23], standard scaling, and Yeo-Johnson transformation [24] as part of the pre-processing pipeline. On the other side, there are LightGBM, XGBoost, and CatBoost that deliver higher accuracy without scaling or normalization. All learners use mean imputation and categorical encoding. However, while XGBoost and LightGBM use m-estimate encoding, CatBoost gives the option to use its built-in categorical encoder, which can enhance performance. Additionally, the ensemble includes three CatBoost models, with two using CatBoost's default Bayesian bootstrapping and one using Bernoulli bootstrapping.

The same cross-validation strategy is used to create an out-of-fold probabilistic prediction, which serves as input for the neural network meta-estimator in the stacked ensemble. In this neural network, all out-of-fold predictions are concatenated and then fed into seven separate fully connected layers with softmax activation function each, and element-wise multiplication is applied between the original model predictions and the softmax outputs. All the multiplication results are added together to create the final probabilistic prediction, and the argmax function is used to decide the final class prediction.

4. Conclusion

This paper presents the two leading solutions from the Cool, Quiet City Kaggle competition of 2023 to 2024, which aimed to predict human comfort in urban environments. The leading models achieved over 0.7 accuracy on multiclass tasks, demonstrating strong performance across multiple comfort-related objectives. Successful approaches consistently used weather, time, and physiological characteristics, highlighting their importance in understanding comfort responses. To improve generalization and reduce overfitting, participants employed model ensembling and multi-fold cross-validation. Several differences between the top solutions were uncovered, including the preprocessing approaches and the structure and components of the machine learning strategy. The competition and resulting solutions are limited in their generalizability in other types of subjective prediction of human preferences in other locations, climates, and objectives. Tree-based models proved especially effective, outperforming others for their suitability for structured and tabular data. This paper documents the setup of the competition, highlights key modeling strategies, and offers a benchmark for future data-driven urban comfort research.

Acknowledgments

This research was supported by the Singapore Ministry of Education Academic Research Fund Tier 1 Grants titled *Ecological Momentary Assessment (EMA) for Built Environment Research* and *Multi-scale Digital Twins for the Urban Environment: From Heartbeats to Cities.* Individual coauthors are also supported by a NUS Graduate Research Scholarship or through the National Research Foundation, the Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

References

[1] Graham L t, Parkinson T and Schiavon S 2021 *Buildings and Cities* **2** 166–184
[2] Miller C, Abdelrahman M, Chong A, Biljecki F, Quintana M, Frei M, Chew M and Wong D 2021 *J. Phys. Conf. Ser.* **2042** 012041
[3] Miller C 2019 *Machine Learning and Knowledge Extraction* **1** 974–993
[4] Fu C, Arjunan P and Miller C 2022 Trimming outliers using trees: winning solution of the large-scale energy anomaly detection (LEAD) competition *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* BuildSys '22 (New York, NY, USA: Association for Computing Machinery) pp 456–461
[5] Miller C, Hao L and Fu C 2022 Gradient boosting machines and careful pre-processing work best: ASHRAE great energy predictor III lessons learned *ASHRAE Transactions* vol 128 (ASHRAE) pp Pages 405–413
[6] Miller C, Chua Y X, Frei M and Quintana M 2022 Towards smartwatch-driven just-in-time adaptive interventions (JITAI) for building occupants *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* BuildSys '22 (New York, NY, USA: Association for Computing Machinery) pp 336–339
[7] Tartarini F, Frei M, Schiavon S, Chua Y X and Miller C 2023 *J. Phys. Conf. Ser.* **2600** 142003
[8] Miller C, Quintana M, Frei M, Chua Y X, Fu C, Picchetti B, Yap W, Chong A and Biljecki F 2023 Introducing the cool, quiet city competition: Predicting smartwatch-reported heat and noise with digital twin metrics *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* BuildSys '23 (New York, NY, USA: Association for Computing Machinery) pp 298–299
[9] Frei M Cozie apple github repository - an ios application for watch surveys and physiological data collection URL `https://cozie-apple.app/`
[10] Yap W, Stouffs R and Biljecki F 2023 *npj Urban Sustain* **3** 1–11
[11] Yap W Urbanity python package github repository - a network-based python package to understand and model urban complexity URL `https://github.com/winstonyym/urbanity`
[12] Miller C Cool, quiet city competition on kaggle - predicting smartwatch-reported heat and noise with digital twin metrics URL `https://www.kaggle.com/competitions/cool-quiet-city-competition`
[13] Ibrahim M Michael ibrahim kaggle profile URL `https://www.kaggle.com/michaelibrahim`
[14] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q and Liu T Y 2017 *Advances in neural information processing systems* **30**

[15] Chen T and Guestrin C 2016 Xgboost: A scalable tree boosting system *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* pp 785–794
[16] Prokhorenkova L, Gusev G, Vorobev A, Dorogush A V and Gulin A 2018 *Advances in neural information processing systems* **31**
[17] Ibrahim M Michael ibrahim first-place solution kaggle notebook URL `https://www.kaggle.com/code/michaelibrahim/1st-place-solution-cool-quiet-city-competition`
[18] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V *et al.* 2011 *the Journal of machine Learning research* **12** 2825–2830
[19] Zhou Z H 2025 *Ensemble methods: foundations and algorithms* (CRC press)
[20] Syah Akbar I Iqbal syah akbar kaggle profile URL `https://www.kaggle.com/iqbalsyahakbar`
[21] Syah Akbar I Iqbal syah akbar second-place solution kaggle notebook URL `https://www.kaggle.com/code/iqbalsyahakbar/2nd-place-cool-quiet-city-competition`
[22] Sechidis K, Tsoumakas G and Vlavahas I 2011 On the stratification of multi-label data *Machine learning and knowledge discovery in databases, part III* Lecture notes on computer science (Springer Science+Business Media) pp 145–158
[23] Micci-Barreca D 2001 *ACM SIGKDD Explorations Newsletter* **3**(1) 27–32
[24] Yeo I K and Johnson R A 2000 *Biometrika* **87**(4) 954–959