ICT for Health & Welfare: Indoor Particulate Matter Does Matter -

Edge-Based Computing for IoT Enabled Indoor Air Quality Monitoring

Assoc.Prof.Dr Thinagaran Perumal, Universiti Putra Malaysia,
Email: thinagaran@upm.edu.my

Background:

A room with more people often has a **lower air quality index**. This is due to a larger crowd being a source of **CO2 and PM**, which worsens air quality in an enclosed space (G. Ciceri et al., 2020). **Clean air** is an important part of **Quality of Life (QoL)** where many researchers strive to improve upon.

Staying indoors for **extended periods** (with poor air quality) can lead to:
- 25% higher risk of lung cancer.
- Acute respiratory issues
- Cardiorespiratory issues
(Eliani et al., 2020)

ICT for Health & Welfare: Indoor Particulate Matter Does Matter:
Edge-Based Computing for IoT Enabled Indoor Air Quality Monitoring

Research aim:

The proposed project aims to enhance and evaluate a machine learning model deployed on the edge to estimate headcount of more and less than seven from air quality data analysis.

Why the need for this study?

The researchers argue that there is a need for an improved headcount **estimator** compared to classification techniques used as a better means to accurately infer headcount of more (or even less) than seven based on new environmental data deployed on the **edge** in an **Indoor Air Quality (IAQ) system** to address the problem.

ICT for Health & Welfare: Indoor Particulate Matter Does Matter:
Edge-Based Computing for IoT Enabled Indoor Air Quality Monitoring

Research comparison:

Performance metrics			
Models	**Acc.(%)**	**MSE**	**Hyper-parameters**
MLP	56.4	0.127	$Optmiz.(Adam)$, $Act.\ Func.(ReLU)$, $HL(4)$, $Neurons([25,20,15,10])$, $Regular.(0.01)$ $LR(0.001)$, $Loss(categ.\ crossentropy)$, $Epochs(500)$, $Output\ function(softmax)$
RF	82.3	0.454	$Rand.\ state(0)$, $Criter.(gini)$, $Max\ depth(2)$, $Max\ features(auto)$, $Estimators(100)$
SVM	96.1	0.235	$Kernel(rbf)$, $C(1.0)$, $Y(auto)$
KNN	97.3	0.037	$N.\ neighbors(1)$, $Weights(uniform)$, $Alg.(auto)$, $Metric(minkowski)$, $n\_jobs(1)$

Table 1: Performance metrics (G. Ciceri et al., 2020)

Previous work concluded with KNN as a preferred model for deployedment at the edge.

A reimplementation and enhancement to KNN was first done.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	K Neighbors Classifier	0.9963	0	0.9963	1.0000	0.9981	0.0000	0.0000

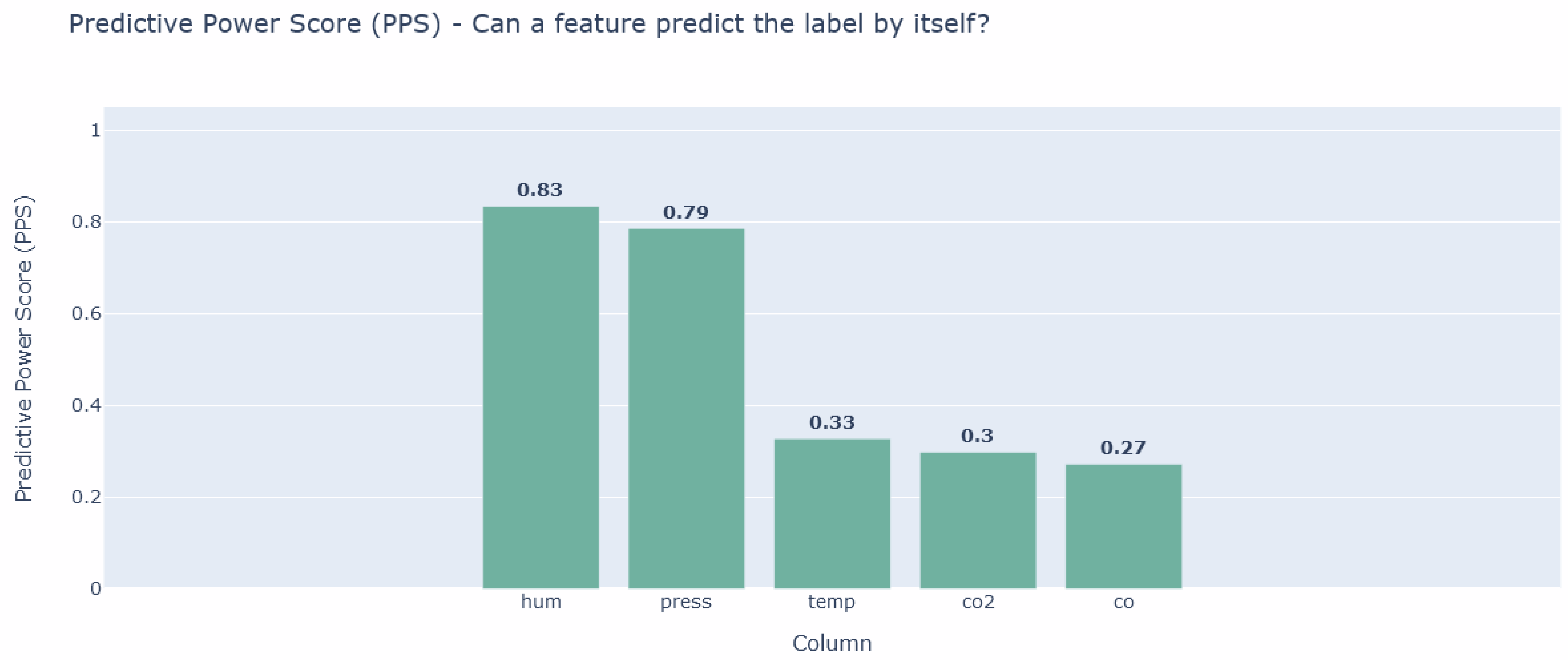
	pm1	temp	hum	press	co	co2	pm2.5	pm10	label	prediction_label	prediction_score
0	6.0	22.5	53.099998	1022.882812	6.00741	486.399658	7.0	7.4	2	2	1.0
1	4.0	22.4	53.200001	1022.958374	5.94448	420.833771	6.0	7.4	2	2	1.0

Figure 1: Reimplementation results with enhancement to KNN classification algo.

Performance metrics			
Models	**Acc.(%)**		**Hyper-parameters**
KNN	99.63		*KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='manhattan', metric_params=None, n_jobs=-1, n_neighbors=13, p=2, weights='distance'))]*

Table 2: Enhanced reimplementation hyperparameters.

Research comparison:



In this case, it is not attributed to data leakage.

Some analysis done shows that while **CO_2** was a primary feature mentioned by the **previous** authors, **humidity** and **pressure** proved to have a higher PPS . Also primary feature in models tested **(both classification or regression-based models.**

ASEAN IVO

ICT for Health & Welfare: Indoor Particulate Matter Does Matter:
Edge-Based Computing for IoT Enabled Indoor Air Quality Monitoring

Treating it as a regression problem:

Largest over estimation errors:

	pm1	temp	hum	press	co	co2	pm2.5	pm10	label	predicted label	label Prediction Difference
958	6.00	22.30	52.00	1022.62	5.56	328.40	10.00	11.40	2	1.94	0.06
963	5.00	21.70	56.70	1020.39	5.37	620.57	6.00	9.40	7	6.96	0.04
735	5.00	21.60	56.70	1020.39	5.41	704.57	6.00	6.40	7	6.97	0.03

Largest under estimation errors:

	pm1	temp	hum	press	co	co2	pm2.5	pm10	label	predicted label	label Prediction Difference
1217	1.30	22.30	58.10	1024.49	6.10	620.57	1.80	9.90	4	4.14	-0.14
1293	6.00	21.90	59.20	1024.99	5.63	670.13	8.00	10.40	4	4.13	-0.13
1194	6.00	21.80	51.90	1022.49	5.44	297.50	7.00	9.40	1	1.08	-0.08

Figure 2: Estimation Tests done on predictions of occupants (label)

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Voting Regressor	0.0049	0.0007	0.0255	0.9999	0.0044	0.0014

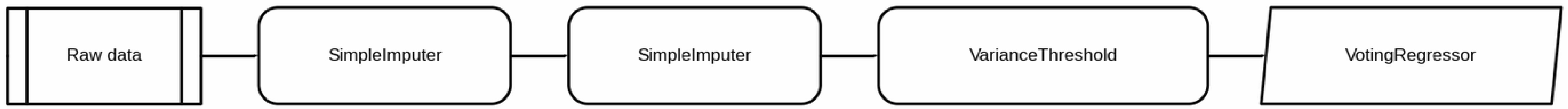


Figure 3: Revision 1 of model performance & Pipeline used

Summary & Conclusion:

Average model inference time for one sample (in seconds): 0.00025608

Deployment on the edge enables quick feedback towards headcount inference.

While occupancy can be a **subjective matter (headcount vs occupancy), Many research papers perceive occupancy as a binary problem (yes/no) and headcount is not needed.** While an **exact** number of headcount is the '**best**'. **But I think a whole number representation is not needed.** Systems can make do with features such as crowd control using an estimated number of people.

Eg:
Actual: 10 people
Predicted 9.5 people

With a slight improvement and synthetic test data to simulate data for more than seven occupants, this project could improve upon the existing headcount implementation with a greater degree of **correctness** and **accuracy**.

Regression models shown will be deployed on a similarly implemented project part of **PutraAIR**. This will make **inferring** based on air quality data an added feature capable of managing situations such as overcrowding in a small space or even detect problems with ventilation (if any).

Nevertheless, the proposed research aims, which are to enhance and evaluate a machine learning model to estimate headcounts of more and less than seven from air quality data analysis is a success.

References:

- Cicceri, G., Scaffidi, C., Benomar, Z., Distefano, S., Puliafito, A., Tricomi, G., & Merlino, G. (2020). Smart Healthy Intelligent Room: Headcount through Air Quality Monitoring. 2020 IEEE International Conference on Smart Computing (SMARTCOMP), 320–325. https://doi.org/10.1109/SMARTCOMP50058.2020.00071

- Ezani, E., Brimblecombe, P., Asha'ari, Z. H., Fazil, A. A., Ismail, S. N. S., Ramly, Z. T. A., & Khan, M. F. (2021). Indoor and Outdoor Exposure to PM2.5 during COVID-19 Lockdown in Suburban Malaysia. Aerosol and Air Quality Research, 21(3), 200476. https://doi.org/10.4209/aaqr.2020.07.0476