Room Occupancy Detection
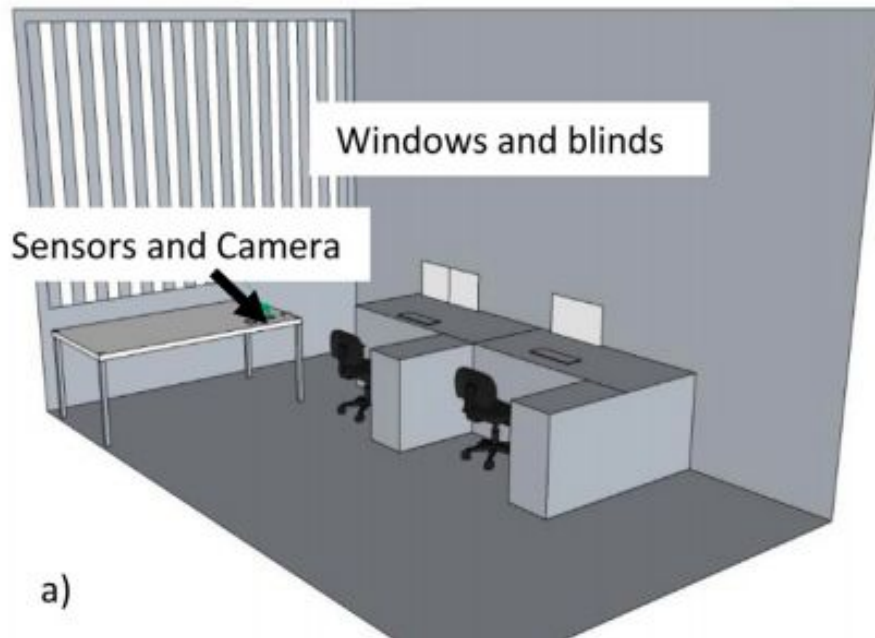
Muhammet Abdullah Soytürk

10.01.2020

❖ Occupancy detection in buildings estimated to **save energy** in the order of 30 to 42%

❖ Detecting the presence of the occupants without using a camera is very interesting due to **privacy concerns**.

Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models. Luis M. Candanedo, VÃ©ronique Feldheim. Energy and Buildings. Volume 112, 15 January 2016, Pages 28-39.
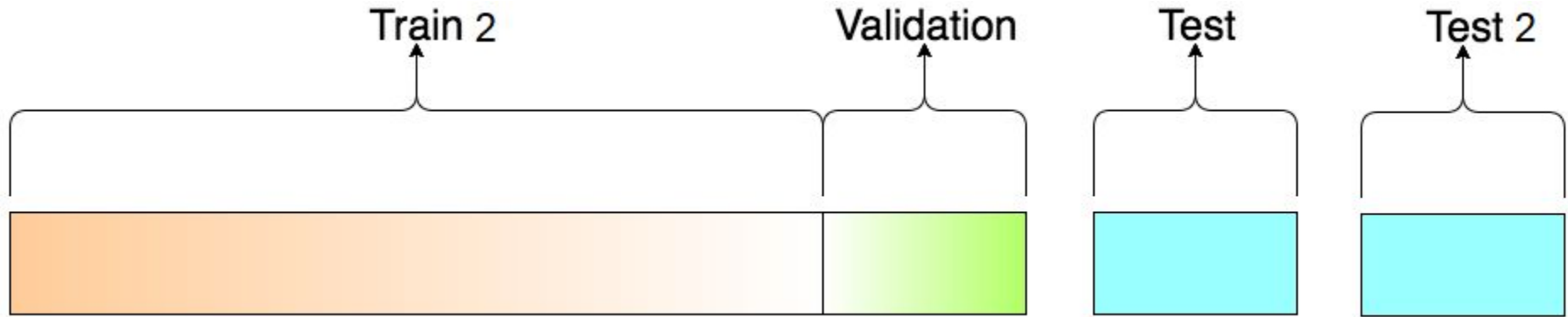
Data set	Number of observations	Data Class Distribution (%)	
		0 (non-occupied)	1 (occupied)
Training	8143 of 7 variables	0.79	0.21
Testing 1	2665 of 7 variables	0.64	0.36
Testing 2	9752 of 7 variables	0.79	0.21

❖ Variables:
  ➢ Temperature (T)
  ➢ Relative Humidity (W)
  ➢ Light
  ➢ CO_2
  ➢ Humidity Ratio (φ)
  ➢ NSM (Substract number of seconds from midnight)
  ➢ Week Status (Weekday or Weekend)

❖ Train_2 -> (6514, 7)

❖ Validation -> (1629, 7)

❖ Train 2, Validation and Test Sets are normalized with MinMax scaling.

	Unnormalized	**Normalized**
Accuracy	0.9941	0.9945

❖ Got a slightly better accuracy.

10-Fold Experiment for Various Classifiers (Q5)

	Accuracy	F-Score	Sensitivity	Specificity	Precision	ROC	Precision Recall Curve	MCC
Logistic Regression	0.9889	0.9746	0.9978	0.9865	0.9526	0.9921	0.9510	0.9680
K Nearest Neighbor (k=1)	0.9937	0.9850	0.9812	0.9970	0.9891	0.9745	0.9745	0.9811
Naive Bayes	0.9894	0.9756	0.9978	0.9871	0.9546	0.9924	0.9529	0.9693
Decision Tree	0.9910	0.9790	0.9790	0.9943	0.9792	0.9867	0.9630	0.9734
Random Forest (estimators=100)	**0.9944**	0.9869	0.9877	0.9962	0.9863	0.9920	0.9768	0.9835
SVM (kern=rbf, C=1, γ=0.125)	0.9887	0.9743	0.9985	0.9861	0.9513	0.9923	0.9502	0.9676
LDA	0.9887	0.9743	0.9985	0.9861	0.9513	0.9923	0.9502	0.9676
AdaBoost	0.9909	0.9790	0.9920	0.9906	0.9665	0.9913	0.9605	0.9734

Feature Selection (Q6)

	Select K Best with f_classif	Select K Best with chi2	Variance Threshold
Logistic Regression	0.9889 (k=5)	0.9889 (k=5)	0.9889
K Nearest Neighbor (k=1)	0.9937 (k=7)	0.9933 (k=7)	0.9937
Naive Bayes	0.9780 (k=3)	0.9894 (k=3)	0.9894
Decision Tree	0.9910 (k=3)	0.9937 (k=3)	0.9912
Random Forest (estimators=100)	0.9932 (k=4)	**0.9953 (k=4)**	0.9944
SVM (kern=rbf, C=1, γ=0.125)	0.9887 (k=6)	0.9887 (k=4)	0.9887
LDA	0.9887 (k=4)	0.9887 (k=4)	0.9887
AdaBoost	0.9909 (k=6)	0.9906 (k=6)	0.9909

I got improvement when I applied feature selection.SelectKBest with chi2 gives the best accuracy.

Hyperparameter Optimization (Q7)

	Optimum Parameters	Accuracies
K Nearest Neighbour	k=1	0.9927
Random Forest	n_estimators=300	**0.9937**
AdaBoost	n_estimators=50	0.9923
Support Vector Machine	C=32, γ=32	0.9933

Accuracies on Validation Set (Q7)

	Accuracy	F-Score	Sensitivity	Specificity	Precision	ROC	Precision Recall Curve	MCC
K Nearest Neighbor (k=1)	**0.9938**	0.9856	0.9942	0.9937	0.9772	0.9939	0.9728	0.9818
Random Forest (estimators=300)	**0.9938**	0.9857	0.9971	0.9929	0.9745	0.9950	0.9723	0.9819
SVM (kern=rbf, C=32, γ=32)	0.9926	0.9829	0.9971	0.9914	0.9691	0.9942	0.9669	0.9883
AdaBoost (estimators=50)	0.9920	0.9814	0.9942	0.9914	0.9690	0.9928	0.9646	0.9765

Stacking Ensemble (Q8)

	Accuracy	F-Score	Sensitivity	Specificity	Precision	ROC	Precision Recall Curve	MCC
Scores with Cross Validation on optimum train two	0.9938	0.9855	0.9848	0.9962	0.9863	0.9905	0.9745	0.9816
Scores on Validation Set	0.9938	0.9857	0.9971	0.9929	0.9745	0.9950	0.9723	0.9819
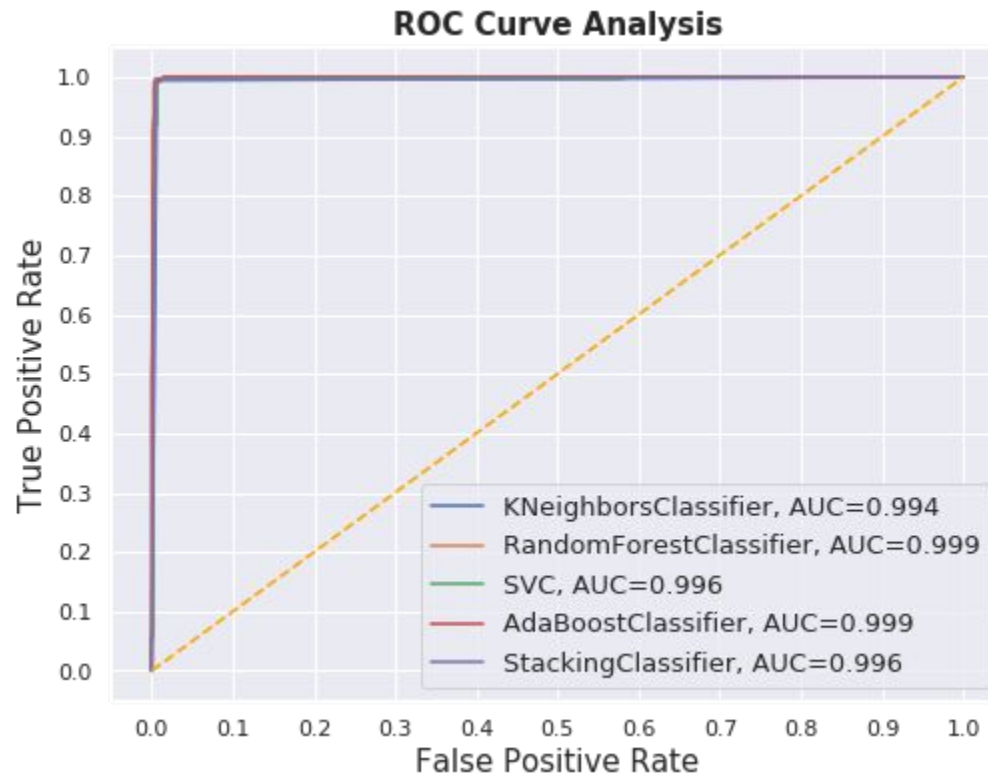
- ❖ Random Forest and AdaBoost performed the best.
- ❖ Since the accuracies are already very high, there is no need to suggest other methods to improve the accuracy.

Accuracies of Best Classifier on Test Sets (Q10)

	Accuracy	F-Score	Sensitivity	Specificity	Precision	ROC	MCC
Test Set 1	0.9459	0.9243	0.9053	0.9692	0.9442	0.9373	0.8828
Test Set 2	0.9735	0.9391	0.9712	0.9741	0.9090	0.9726	0.9230

Literature Review & Comparison (Q11)

Model	Parameters	Training Accuracy	Test 1 Accuracy	Test 2 Accuracy
Random Forest	T, ϕ, Light, CO_2, W	100	95.53	98.06
GBM	Light	98.77	97.86	99.32
CART	Light	98.78	97.86	99.31
LDA	T, ϕ, Light, CO2, W, NS, WS	98.85	**97.90**	**99.33**