Human Occupancy Detection Using Smart Energy Meter.

Presented By -

Rohit Chavan

Github -

github.com/rdchavan/Capstone-Project-CS668

Research Question

- Can one predict if a certain household in Zurich is currently occupied or not just by looking at their energy consumption and changes in the activation state of appliances?
- Can the model be further improved to predict occupancy if new feature set are considered?

Motivation

I am interested in working on this data set because it presents a special challenge for determining a person's daily energy use and how this knowledge can be utilized to further reform public policy choices about energy use, production, and distribution, assisting in making sure that resources are used as effectively and efficiently as possible.

Dataset

- **Smart meter.zip-** 1 Hz aggregate consumption data. Each measurement contains data on current, voltage, and phase shift for each of the three phases in the household.
- **Plug Data.zip -** 1 Hz plug-level data measured from selected appliances.
- **Occupancy Data.csv -** Occupancy information measured through a tablet computer (manual labeling) and a passive infrared sensor (in some of the households).
- **Measurement period:** 27.06.12 to 23.01.13

```
df_sm.head(5)
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	28.337	14.774	8.221	5.342	0.0768	0.1525	0.1130	0.0392	240.06	241	241	240	120	298	287	308
1	28.987	15.104	8.461	5.422	0.0772	0.1536	0.1133	0.0398	239.94	241	241	240	120	299	288	307
2	29.061	15.215	8.409	5.437	0.0771	0.1543	0.1130	0.0404	239.94	241	240	240	120	299	288	307
3	28.624	14.881	8.325	5.418	0.0783	0.1532	0.1127	0.0403	239.83	241	241	240	120	298	288	307
4	28.604	14.913	8.245	5.446	0.0752	0.1537	0.1123	0.0417	239.83	241	241	240	120	298	287	307

```
df_occ.head(5)
```

	Unnamed: 0	'00:00:00'	'00:00:01'	'00:00:02'	'00:00:03'
0	15-Jul-2012	1	1	1	1
1	16-Jul-2012	1	1	1	1
2	17-Jul-2012	1	1	1	1
3	18-Jul-2012	1	1	1	1
4	20-Jul-2012	1	1	1	1

5 rows × 86401 columns

```
df_plug.head(5)
```

	0
0	49.2516
1	49.2516
2	49.2516
3	51.3899
4	49.2516

Dataset contd.

```
[137] pd.read_csv('drive/MyDrive/Capstone/Dataset/cleaned_data_1m.csv')
```

	Datetime	p_total	p_phase1	p_phase2	p_phase3	i_neutral	i_phase1	i_phase2	i_phase3	v_phase1	...	app_03	app_04	app_05	app_06
0	2012-06-01 00:00:00	1020.0	100.0	70.0	850.0	3.0	1.0	0.0	4.0	238.0	...	0.0	0.0	5.0	76.0
1	2012-06-01 00:01:00	1018.0	101.0	69.0	848.0	3.0	1.0	0.0	4.0	237.0	...	0.0	0.0	5.0	73.0
2	2012-06-01 00:02:00	1031.0	120.0	64.0	848.0	3.0	1.0	0.0	4.0	236.0	...	0.0	0.0	5.0	71.0

```
pd.read_csv('drive/MyDrive/Capstone/Dataset/cleaned_data_1m.csv').shape()
```

```
(1766880, 31)
```

```
df.shape
```

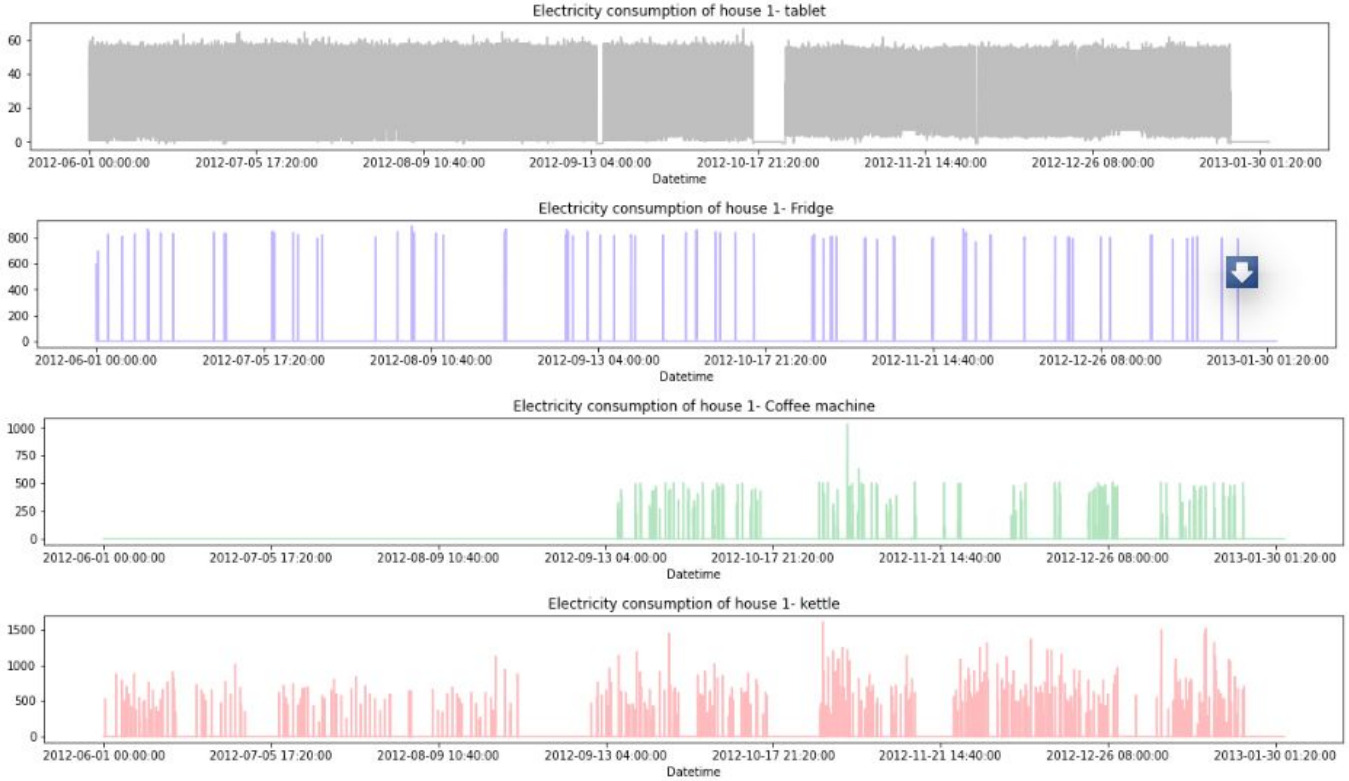
```
(353376, 30)
```

Literature Review Summary

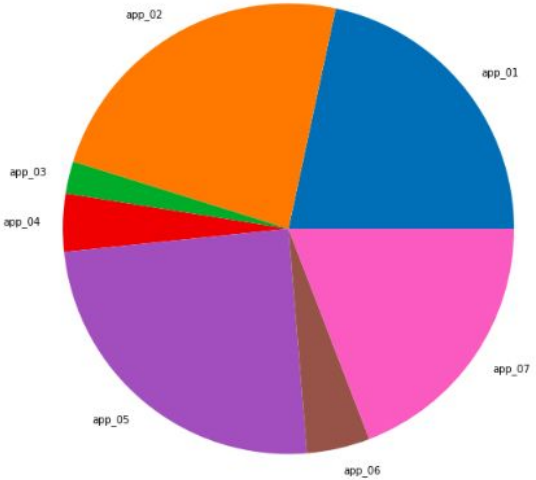
- The location and kind of appliances among other variables were assessed for their effects on the accuracy of the model wherein living room appliance and kitchen appliance played important role.
- Model with convolutional Neural network achieved a better performance than KNN and Random forest model throughout various seasons and days of the week and was resistant to various occupancy patterns.
- Using a combination of Gaussian Mixture Model (GMM) and Unsupervised Anomaly Detection the author was successful in predicting human occupancy without the use of training data.
- An ADOBE-Net model achieved a better accuracy than traditional ML model on multiple public dataset. Also the model further be enhanced by adding more attention layers.
- An adversarial machine learning approach utilizing Generative Adversarial Network (GANs) may successfully safeguard people's privacy from smart meter data, which is crucial for guaranteeing data security and privacy in the context of smart grids and energy management systems.

.

EDA

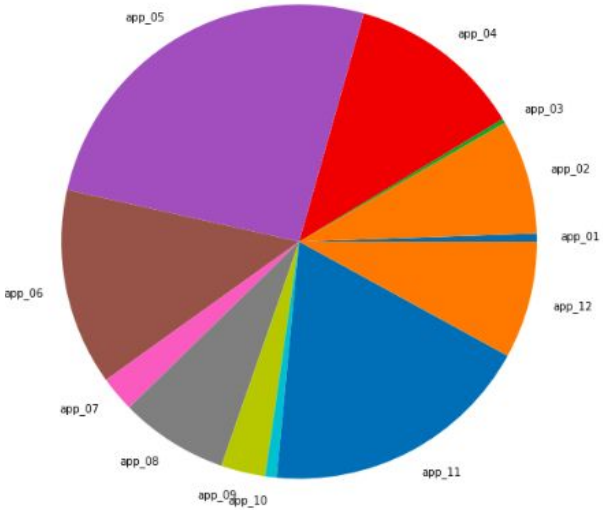

Electricity consumption of house 1- tablet
60
40
20
0
2012-06-01 00:00:00
2012-07-05 17:20:00
2012-08-09 10:40:00
2012-09-13 04:00:00
2012-10-17 21:20:00
2012-11-21 14:40:00
2012-12-26 08:00:00
2013-01-30 01:20:00
Datetime
Electricity consumption of house 1- Fridge
800
600
400
200
0
2012-06-01 00:00:00
2012-07-05 17:20:00
2012-08-09 10:40:00
2012-09-13 04:00:00
2012-10-17 21:20:00
2012-11-21 14:40:00
2012-12-26 08:00:00
2013-01-30 01:20:00
Datetime
Electricity consumption of house 1- Coffee machine
1000
750
500
250
0
2012-06-01 00:00:00
2012-07-05 17:20:00
2012-08-09 10:40:00
2012-09-13 04:00:00
2012-10-17 21:20:00
2012-11-21 14:40:00
2012-12-26 08:00:00
2013-01-30 01:20:00
Datetime
Electricity consumption of house 1- kettle
1500
1000
500
0
2012-06-01 00:00:00
2012-07-05 17:20:00
2012-08-09 10:40:00
2012-09-13 04:00:00
2012-10-17 21:20:00
2012-11-21 14:40:00
2012-12-26 08:00:00
2013-01-30 01:20:00
Datetime

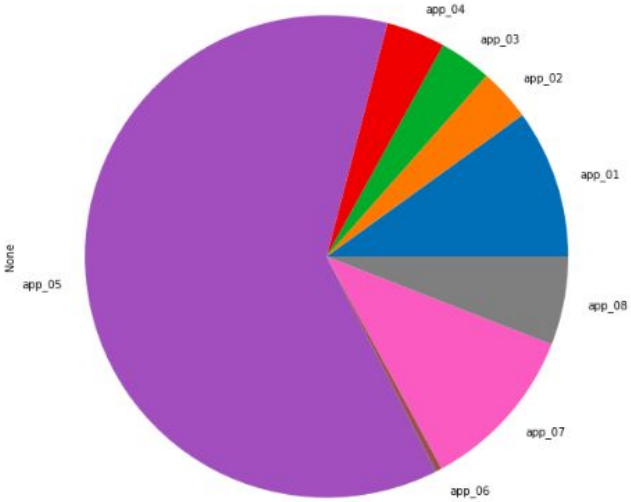
EDA



Appliance like washer,fridge ,dryer, freezer are widely used for household 1

Entertainment Appliance, TV, Freezer are widely used for household 2





Appliance like Freezer, Tablet, Lamp are widely used for household 4

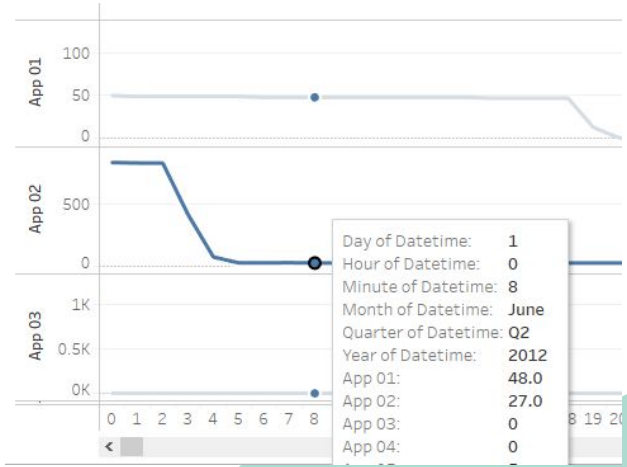
Methodology

- Consolidate and resampled from sec to 1 min all the individual file related to Smart meter reading, Plug and occupancy file into single master file based on household and Datetime variable. (Fig 1)
- Finding the transient value of each household appliance as some might consume energy when idle and use it to considered on /off state of the appliance. (Fig 2)
- As the target variable occupancy_tablet has binomial value thus model such as Decision tree, Random Forest, Logistic Regression, Gradient boosting and neural network can be applied.
- The model are then evaluated based on Accuracy, precision, recall, F1 Score and AUC curve. Further dataset can be improved by adding public Holiday and testing the effect on the model.

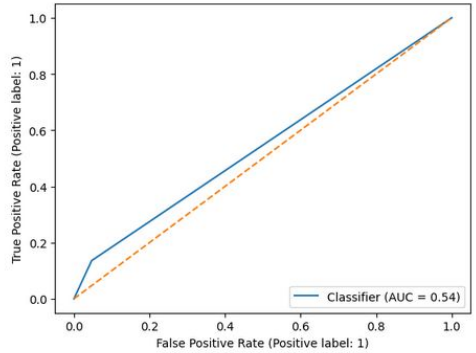
Fig 1

Datetime	p_total	app_01	app_03	app_04	app_05	app_06	app_07
2012-06-01 00:00:00	1020.0	1.0	0.0	0.0	0.0	1.0	0.0
2012-06-01 00:01:00	1018.0	1.0	0.0	0.0	0.0	1.0	1.0
2012-06-01 00:02:00	1031.0	1.0	0.0	0.0	0.0	1.0	1.0
2012-06-01 00:03:00	638.0	1.0	0.0	0.0	0.0	1.0	1.0
2012-06-01 00:04:00	279.0	1.0	0.0	0.0	0.0	1.0	1.0
...	...	...	...	...	...	...	...
2013-01-31 23:55:00	375.0	1.0	0.0	0.0	1.0	0.0	0.0
2013-01-31 23:56:00	374.0	1.0	0.0	0.0	1.0	0.0	0.0
2013-01-31 23:57:00	372.0	1.0	0.0	0.0	1.0	0.0	0.0
2013-01-31 23:58:00	370.0	1.0	0.0	0.0	1.0	0.0	0.0
2013-01-31 23:59:00	617.0	1.0	0.0	0.0	1.0	0.0	0.0

Fig 2

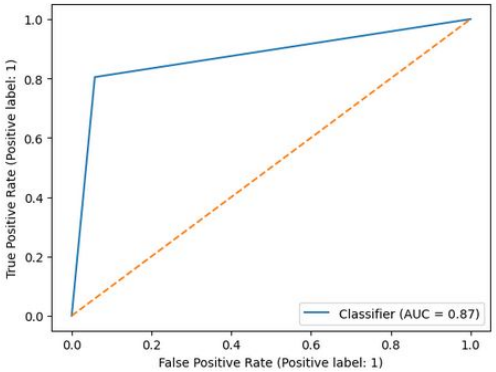


Experiment (Raw Data)

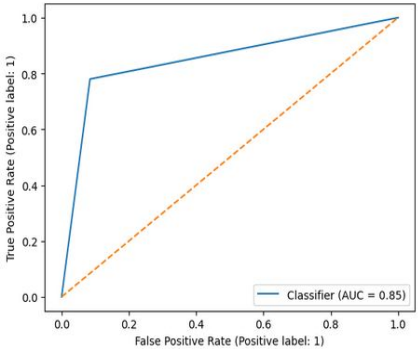


AUC : 0.544409

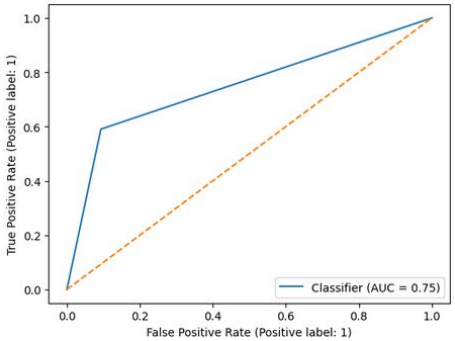
Linear Regression



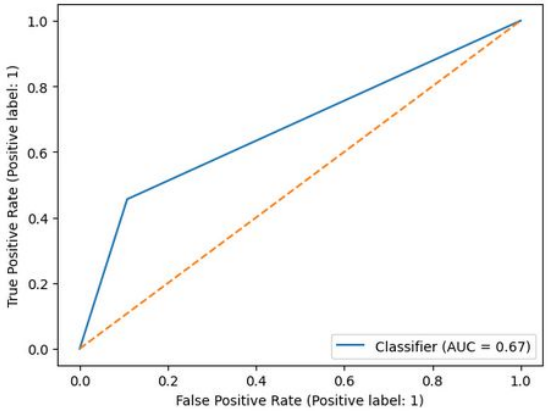
Random Forest



Decision Tree

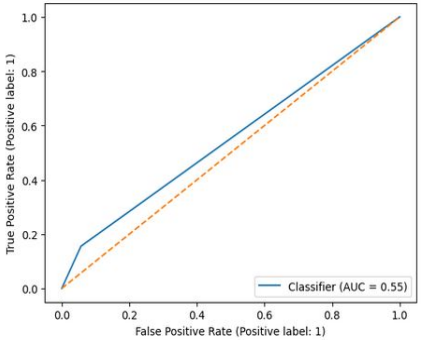


Gradient Boosting

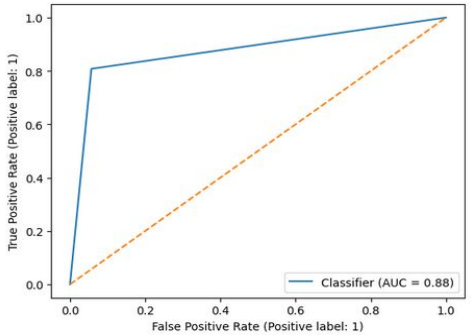


Neural Network

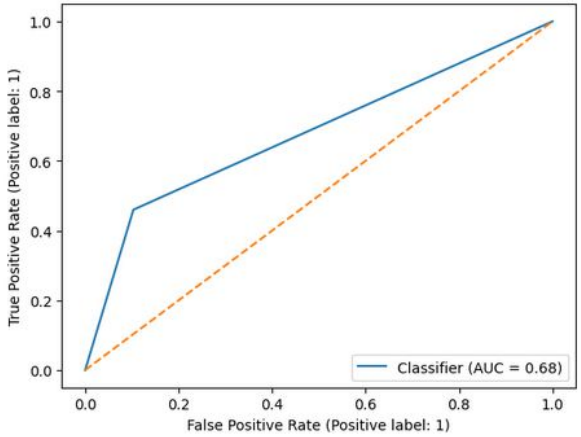
Experiment (Raw Data + Public Holiday Data set)



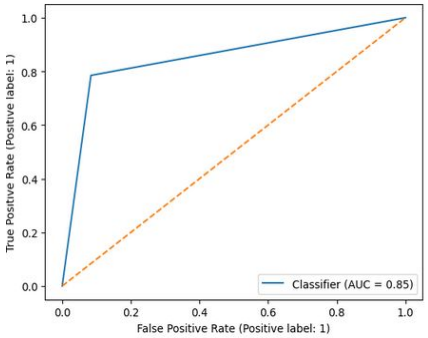
Linear Regression



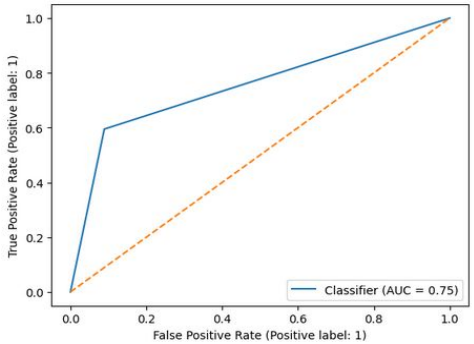
Random Forest



Neural Network



Decision Tree



Gradient Boosting

Experiment outcome

Metric	Raw Dataset					Raw Data + Public Holiday Dataset				
	LR	DT	RF	GB	NN	LR	DT	RF	GB	NN
Accuracy	67%	87%	89%	80%	74%	67%	87%	90%	80%	75%
Precision	60%	83%	88%	77%	69%	59%	83%	88%	78%	70%
Recall	14%	78%	80%	59%	46%	16%	78%	81%	59%	46%
F1 Score	22%	80%	84%	67%	55%	25%	81%	84%	67%	56%
AUC	0.544	0.847	0.873	0.748	0.674	0.549	0.850	0.8755	0.752	0.678