|  |  |  |
| --- | --- | --- |
| **PROJECT OVERVIEW** | **Project Name: Human occupancy detection using Smart energy meter**. | **Project Manager:** |

**STATEMENT (POS)**

*Maximum 4 lines / section*

**Problem/Opportunity/Research Question(s):**

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?

**Dataset:**

The ECO (Electricity Consumption and Occupancy) data set is a comprehensive open-source (Creative Commons License CC BY 4.0) data set for non-intrusive load monitoring and occupancy detection research. It was collected in 6 Swiss households over a period of 8 months. For each of the households, the ECO data set provides:

* 1 Hz aggregate consumption data. Each measurement contains data on current, voltage, and phase shift for each of the three phases in the household.
* 1 Hz plug-level data measured from selected appliances.
* Occupancy information measured through a tablet computer (manual labeling) and a passive infrared sensor (in some of the households).

**SMART METER DATA:**

Measurement period:

01.06.12 to 31.01.13

Coverage:

No. days: 245, Coverage: 99.64%

Limitation:

The first 29 days (i.e., from 01.06.12 to 29.06.12), each of the power measurements has been rounded to 10W precision.

Description:

Each CVS file contains aggregate consumption data of a single day and is named accordingly (e.g., "2012-08-31.csv"). Each file contains 86,400 rows (i.e., one row for each second). Rows with missing measurements are denoted as "-1". The coverage specifies the proportion of "existing" values, i.e., values that are > -1.

The rows consist of the following comma separated values:

* powerallphases: Sum of real power over all phases
* powerl1: Real power phase 1
* powerl2: Real power phase 2
* powerl3: Real power phase 3
* currentneutral: Neutral current
* currentl1: Current phase 1
* currentl2: Current phase 2
* currentl3: Current phase 3
* voltagel1: Voltage phase 1
* voltagel2: Voltage phase 2
* voltagel3: Voltage phase 3
* phaseanglevoltagel2l1: Phase shift between voltage on phase 2 and 1
* phaseanglevoltagel3l1: Phase shift between voltage on phase 3 and 1
* phaseanglecurrentvoltagel1: Phase shift between current/voltage on phase 1
* phaseanglecurrentvoltagel2: Phase shift between current/voltage on phase 2
* phaseanglecurrentvoltagel3: Phase shift between current/voltage on phase 3

Matlab files is also available, which contain the same data stored as Matlab structs (using the field names described above). Each Matlab file contains aggregate consumption data of a single day (e.g., "2012-08-31.mat") and contains a single struct. The name of the struct encodes the date, number of household, and type of data ("00" represents a smart meter, "01" represents plug no. 1, ...). For instance, "Appliance010020120831" contains smart meter data from household 01 measured on August 31st, 2012.

**PLUG DATA:**

Like the smart meter data, the plug data is also provided on a daily based (i.e., 86,400 measurements per day) both in CSV and Matlab format. In contrast to the smart meter data, it contains only one value: \* consumption: Real power measured by the plug. Each file contains plug measurements for the whole household. The (daily) data for each plug is stored in subfolders named after the plug ID (e.g., "05") described above.

Measurement period:

01.06.12 to 31.01.13

01: Tablet (no. days: 240, coverage: 97.43%)

02: Dishwasher (no. days: 240, coverage: 97.09%)

03: Air exhaust (no. days: 240, coverage: 96.18%) (\*)

04: Fridge (no. days: 240, coverage: 98%)

05: Entertainment (no. days: 240, coverage: 96.18%) (\*\*)

06: Freezer (no. days: 240, coverage: 96.39%)

07: Kettle (no. days: 240, coverage: 88.5%)

08: Lamp (no. days: 240, coverage: 90.21%) (\*\*\*)

09: Laptops (no. days: 240, coverage: 83.36%)

10: Stove (no. days: 28, coverage: 100%) (\*\*\*\*)

11: TV (no. days: 240, coverage: 100%) (\*\*)

12: Stereo (no. days: 240, coverage: 95.95%) (\*\*)

Comments:

(\*) The air exhaust is located above the stove. We deployed a plug to identify when inhabitants were cooking, because the stove cannot be connected to a smart plug.

(\*\*) The entertainment system consists of a stereo system and a TV. We 'manually' disaggregated those two into plugs 11 and plugs 12. See the helper functions of the NILM-Eval project for more details. For this reason, adding the consumption of plugs 11 and 12 equals the consumption of plug 05. Except during 08.09.2012 and 13.09.2013 - during this time, only the TV was connected to plug 05.

(\*\*\*) This is a dimmable lamp, whose power consumption depends on the setting.

(\*\*\*\*) We extracted the power consumption of the stove 'manually': Therefore, we made use of (1) the consumption of the air exhaust, and (2) the fact that the stove is the only appliance that consumes power on two phases at the same time (which means it is possible to detect switching events by investigating the power consumption measured on different phases at the same time).

**OCCUPANCY DATA**

The occupancy data is available for summer (01\_summer.csv) and winter (01\_winter.csv). Each row in the csv file represents a day (specified by the first column). Each day consists of 86400 fields indicating presence (1) and absence (0) for each second of the day. Thus, the csv file contains 86401 columns.

Occupants specified presence/absence through a tablet computer. As this process is prone to errors (like any method capturing ground truth...) we filtered the data as described in our paper "Occupancy Detection from Electricity Consumption Data". Here we provide the filtered data, which we also used for our analysis to infer occupancy from electricity consumption data.

**Goal:**

Build a model that best fit the data and understanding how accurately the model can predict occupancy of a household and present finding in due date time.

**Objectives:**

* Literature review by researching related paper and accessing way to approach the problem statement.
* Getting familiarize with data by cleaning and performing Explanatory Data Analysis.
* Build a model and analyze the outcome.
* Study the effect of adding new features and its effect on predicting the outcome.
* Compare which option is best suited for the problem statement.

**Success Criteria:**

* Completing the project in due time.
* The model has improved performance than the current performance.
* Presenting all finding and fact.

**Assumptions, Risks, Obstacles:**

* Data quality is not as good higher missing variables or inaccurate measurements.
* Data might be biased or imbalanced.
* Model performance doesn’t improves as expected.

|  |  |  |  |
| --- | --- | --- | --- |
| **Prepared By** | **Date** | **Approved By** | **Date** |
| Rohit Chavan | 02/14/2023 |  |  |