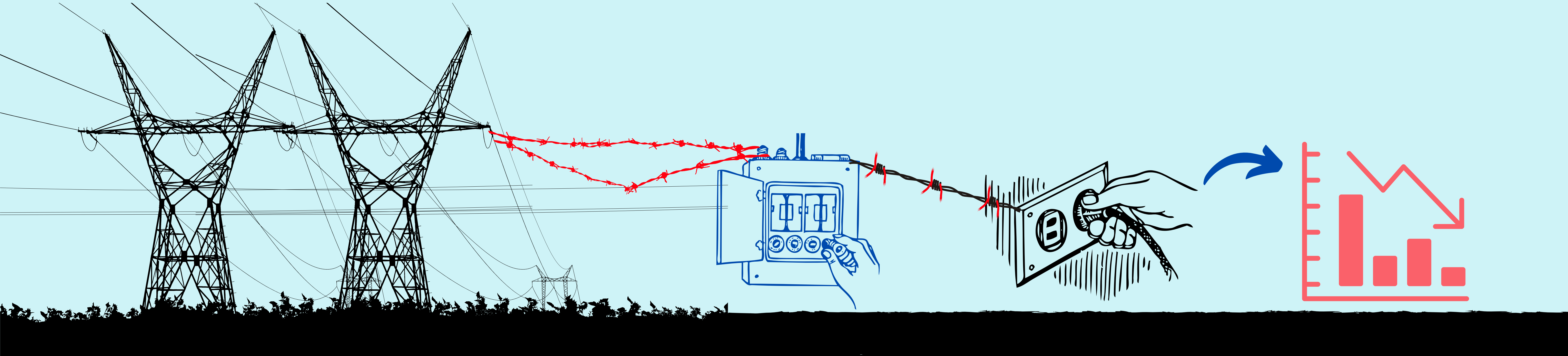
AIML Course Project

Course Instructor: Dr. Manish Chaturvedi

Electricity Theft Detection Using Machine Learning Techniques



Flow Of The Presentation

1. Understanding Electricity Theft & Its Implications
2. Motivation for the Project
3. Dataset Used & Data Dictionary
4. Work Done So Far
5. Type of Problem : Machine Learning Perspective
6. Planned Work

Understanding Electricity Theft & Its Implications

Electricity Losses :

Technical Losses: Occurs in Generation, Transmission & Distribution

Non - Technical Losses: **Electricity Theft**, Conveyance, Unmetered Supplies

Bypassing Meter

Tampering Meter Reading

Hacking Meter

Understanding Electricity Theft & Its Implications

Electricity Losses :

Technical Losses: Occurs in Generation, Transmission & Distribution

Non - Technical Losses: **Electricity Theft**, Conveyance, Unmetered Supplies

Bypassing Meter

Tampering Meter Reading

Hacking Meter

Implications

Electrical Systems

Surging electricity, Heavy load on electrical systems, Fires & Ele. Shocks

Economical

Revenue Losses of Power Company, Inadequate Supply, Poor Service & Higher Maintenance

Motivation for the Project

The Motivation for choosing Electricity Theft Detection for this course project is as :

- Real-World Problem
- Companies face huge losses
- Related to Domain
- Realistic Dataset
- Application of Machine Learning in Industry

Dataset Used & Data Dictionary

The dataset for our project is a **realistic electricity consumption** dataset, released by State Grid Corporation of China.

https://www.sgcc.com.cn/ywlm/index.shtml

	CONS_NO	FLAG	2014/1/1	2014/1/10	2014/1/11	2014/1/12	2014/1/13	2014/1/14	2014/1/15	2014/1/16	2014/1/17	2014/1/18
0	0387DD8A07E07FDA6271170F86AD9151	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	01D6177B5D4FFE0CABA9EF17DAFC2B84	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	4B75AC4F2D8434CFF62DB64D0BB43103	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	B32AC8CC6D5D805AC053557AB05F5343	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	EDFC78B07BA2908B3395C4EB2304665E	1	2.90	3.42	3.81	4.58	3.56	4.25	3.86	3.53	3.41	0.85
5	6BCFD78138BC72A9BA1BFB0B79382192	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	34C1954AA3703C4F8BD8EAEA7C4B7B83	1	0.11	0.53	0.45	0.51	1.32	0.71	0.12	0.52	0.55	0.74
7	768309B0EB11FD436CEE5ABFB84F4C0C	1	0.91	0.86	1.10	0.66	5.82	3.17	1.18	4.05	3.66	3.21
8	D0A186208CE83FBCCF730857C9A75B6F	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	516954F5FF177CE314656D727FCC66A5	1	11.02	8.24	7.94	7.92	8.31	7.39	8.27	8.05	8.95	8.32

10 rows × 1036 columns

CONS_NO : Customer Number (Unique ID)

FLAG: Electricity Theft(1) or NOT(0)

Date: January 1, 2014 to October 31, 2016

Dataset Size: 167 MB

Shape : 42372 Rows, 1036 Columns

Each row corresponding usage for a particular customer

The dataset contains the electricity consumption data of **42,372** electricity customers within a time frame of **1,035** days

Type of Problem : Machine Learning Perspective

	CONS_NO	FLAG	2014/1/1	2014/1/10	2014/1/11	2014/1/12	2014/1/13	2014/1/14	2014/1/15	2014/1/16	2014/1/17	2014/1/18
0	0387DD8A07E07FDA6271170F86AD9151	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	01D6177B5D4FFE0CABA9EF17DAFC2B84	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	4B75AC4F2D8434CFF62DB64D0BB43103	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	B32AC8CC6D5D805AC053557AB05F5343	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	EDFC78B07BA2908B3395C4EB2304665E	1	2.90	3.42	3.81	4.58	3.56	4.25	3.86	3.53	3.41	0.85
5	6BCFD78138BC72A9BA1BFB0B79382192	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	34C1954AA3703C4F8BD8EAEA7C4B7B83	1	0.11	0.53	0.45	0.51	1.32	0.71	0.12	0.52	0.55	0.74
7	768309B0EB11FD436CEE5ABFB84F4C0C	1	0.91	0.86	1.10	0.66	5.82	3.17	1.18	4.05	3.66	3.21
8	D0A186208CE83FBCCF730857C9A75B6F	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	516954F5FF177CE314656D727FCC66A5	1	11.02	8.24	7.94	7.92	8.31	7.39	8.27	8.05	8.95	8.32

10 rows × 1036 columns

Type: Classification (*Is stealing electricity* [1], *is NOT stealing electricity* [0]) | Binary Classification

Classification Algorithms: Logistic Regression, Decision Trees, SVM, Naive Bayes, KNN, Random Forest

Planned Work

	CONS_NO	FLAG	2014/1/1	2014/1/10	2014/1/11	2014/1/12	2014/1/13	2014/1/14	2014/1/15	2014/1/16	2014/1/17	2014/1/18
0	0387DD8A07E07FDA6271170F86AD9151	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	01D6177B5D4FFE0CABA9EF17DAFC2B84	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	4B75AC4F2D8434CFF62DB64D0BB43103	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	B32AC8CC6D5D805AC053557AB05F5343	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	EDFC78B07BA2908B3395C4EB2304665E	1	2.90	3.42	3.81	4.58	3.56	4.25	3.86	3.53	3.41	0.85
5	6BCFD78138BC72A9BA1BFB0B79382192	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	34C1954AA3703C4F8BD8EAEA7C4B7B83	1	0.11	0.53	0.45	0.51	1.32	0.71	0.12	0.52	0.55	0.74
7	768309B0EB11FD436CEE5ABFB84F4C0C	1	0.91	0.86	1.10	0.66	5.82	3.17	1.18	4.05	3.66	3.21
8	D0A186208CE83FBCCF730857C9A75B6F	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	516954F5FF177CE314656D727FCC66A5	1	11.02	8.24	7.94	7.92	8.31	7.39	8.27	8.05	8.95	8.32

10 rows × 1036 columns

Type: Classification (*Is stealing electricity* [1], *is NOT stealing electricity* [0]) | Binary Classification

Classification Algorithms: Logistic Regression, Decision Trees, SVM, Naive Bayes, KNN, Random Forest

Planned Work :

- Find relevant patterns in data to be used by Classification Model for Classification
- Using various ML algorithms for performing classification
- Comparing the results obtained & corresponding analysis of obtained results