POWER COMPANY CASE STUDY

Note for every step documented –
1 – Modelling techniques will be in RED.
2 – Data given will be in BLUE.
3 – End result of application in PURPLE.

PROBLEM DATA –

1. Full Predictor Set – Residence Owners:

→ Legal Name (string)
→ SSN (integer)
→ Address (string, google maps coordinates tuple)
→ Amount Outstanding (float)
→ Income Value (integer)
→ Current Profession Type (string – categorical)
→ Profession End Date (date)
→ Mortgage and Loans Rating (integer)
→ Total Family Income (integer)
→ Payment Trail (binary array)
→ Average Extra Payment (integer)

2. Full Predictor Set – Workers:

→ Employee ID (integer)
→ Worker Days (array of strings)
→ Pay Per Day (integer)
→ Age (string - categorical)
→ Family Supporter Score (integer)

3. Full Predictor Set – Current Monetary State of Power Company:

→ Reduction / Benefit Value from Current Tax Laws and Concessions (integer)
→ Funds Immediately Available (integer)
→ Non-Liquid Funds Availability Amount (integer)
→ Non-Liquid Funds Available Time (date)
→ Cost of Electricity Per House (integer)
→ Record of Funds at Critical Histories (array of arrays)

PROBLEM GOAL –

Find the optimum shutoffs per month considering the tradeoff between –

- *workers time available, distance to shutoffs, current monetary resources to pay employees, underestimated limit of how much excess may be requested from number of payers…*
- *against number of payment misdemeanors to deny electricity to, cost saved by shutting off power for them, and cost loss on second chances not given.*

SOLUTION –

Step 1: Cleaning and Preprocessing

a)
Remove outliers.

b)
For all values missing, perform case-based <simulation> to impute missing values.

c)
Divide data into type I and type II based on given predictor information –

- Income Value
- Profession End Date
- Mortgage and Loans Rating

Use <Soft SVM> to divide into type I – Capability to pay and type II – incapability to pay. We want to use soft SVM since income with other two predictors can affect the payment capacity very much and it's common that quite some people have high income but much higher debts, unreliable jobs, and mortgages.

d) Prepare two secondary divisions – one biased towards false positives (allows more people who PROBABLY can pay but are excused) and the other biased towards false negatives (allows more people who PROBABLY can't pay but are asked to pay).

Secondary classification 1 is used for times of surplus balance, and secondary classification 2 is used in times where deficit needs to be covered urgently.

Step 2: Prediction of Repeat Misdemeanors – For those who can pay!

First using the binary payment trail (months paid in the past), we use <auto-regressive exponential smoothing> to score the probability of payment on current month. Let us call this probability p. This method is used due to likelihood of differences being 1st order and being dependent on record of offenses in the past (someone who does it once is likely to do it again, and the longer the bad habit, the harder it is to break).

Using given data –

- Income Value

- Profession End Date
- Mortgage and Loans Rating
- Current Profession Type
- Amount Outstanding
- Total Family Income
- Probability of payment this month - p

Now we build a <naïve bayes hierarchical classification> model where the classes of importance given the predictor values are whether they are likely to pay the amount over the next months based on past payment data, or not. Bayesian hierarchical classification does not assume an underlying distribution and seeing as we demanded a lot of data and may need to have synthetically generated a lot of it, Bayesian models help stay on the safer side.

From this we calculate an estimate of total incoming supply of payment that will support wages for employees and storage for subsidies for underprivileged users.

Step 3: Resource Management

Let us go backwards here –

1. Choose a cutoff for misdemeanor handling.

2. Build graph for shutoffs for the month.

3. Use a <optimized network model> with the costs, worker days, current company repository wealth, and address data on Google Maps, to build an optimized model to assign routes to every worker to cover these shutoff points.

- Constrained by maximum capacity of total cost is current monetary funds available to pay workers and predicted excess payment from regular bills.
- Minimize total time travelled (and hence cost) and return path for each worker.

If no feasible solution, reduce the misdemeanor cutoff OR remove certain misdemeanor locations strategically and go back to step 2.

Else, progress to step 4.

4. Calculate the remainder of the wealth left after paying employees for the shutoff operations.

5. Based on remainder of wealth and month's end, SVM cutoff, divide required excess amount in that ratio among regularly billed customers and pay for subsidy.

6. If there is an excess of this, increase cutoff of misdemeanors for the next month since you have some leeway to afford the cost of losing customers who might eventually pay.