RAKATHON
Disrupt to Dominate

IIT (ISM) Dhanbad

Pawan Kumar Sahu
Pawan Raj
Pratik Bedre

What problem are we solving?

- In most cases, the promotional content and offers from credit card company through email get **under utilised**, ignored or are simply lost due to flood of information on a consumer's device.

- To use **AI** and other technologies to offer customized and targeted offers to clients which is **easy** as well as **safe** to consume.

What’s the Deal?

Credit card companies has **hundreds of offers** stacked up with **millions of consumers** with their own differences on how they use credit card in their daily life. Companies need to make their consumers aware of their offers may it be some introductory offers or some long term plans, some pharmacy offers or some offers on groceries etc. So, does the company need to mail info about all these offers to **every consumer** in hope they go through all of them and pick ones they find **useful**?

Our solution says **“NO”**. We propose our solution to deal with this problem by bringing **relevancy** in company’s mails and **automate the whole process using AI technology**. Offer mails need not to be a source of headache for consumers and just a formality work for company BUT an important **link** of information between company and its customers which benefit both sides, rather than just being ignored.

Our Solution

Expense-related personalisation:

Personalised offers based on consumer’s expenses and credit card usage as mode of payment.

Sector-specific personalisation:

Personalised offers based on consumer’s spending habits/trends.

Transparent

Secure

Scalable

Content

Overview

Collecting transaction details from credit card providers

Analyzing the transaction patterns and volumes.

FInding the best offers and experience for consumers.

Improving consumer experience via new data and feedback

Customized and targeted offers to clients which is easy as well as safe to consume.

Basic Overview of Steps Involved

Step I: Data Collection

- Credit card transaction **details would be collected** from card providers.
- This would only contain **non-sensitive** information as spent amounts, and the POS destination.

Step II: Data Analysis and AI:

- The collected data would be then **cleaned**, **normalised** and used for **predictive** recommendations.
- Machine learning techniques like data **clustering**, **time-series analysis**, **collaborative filtering** would be employed to improve customer experience.

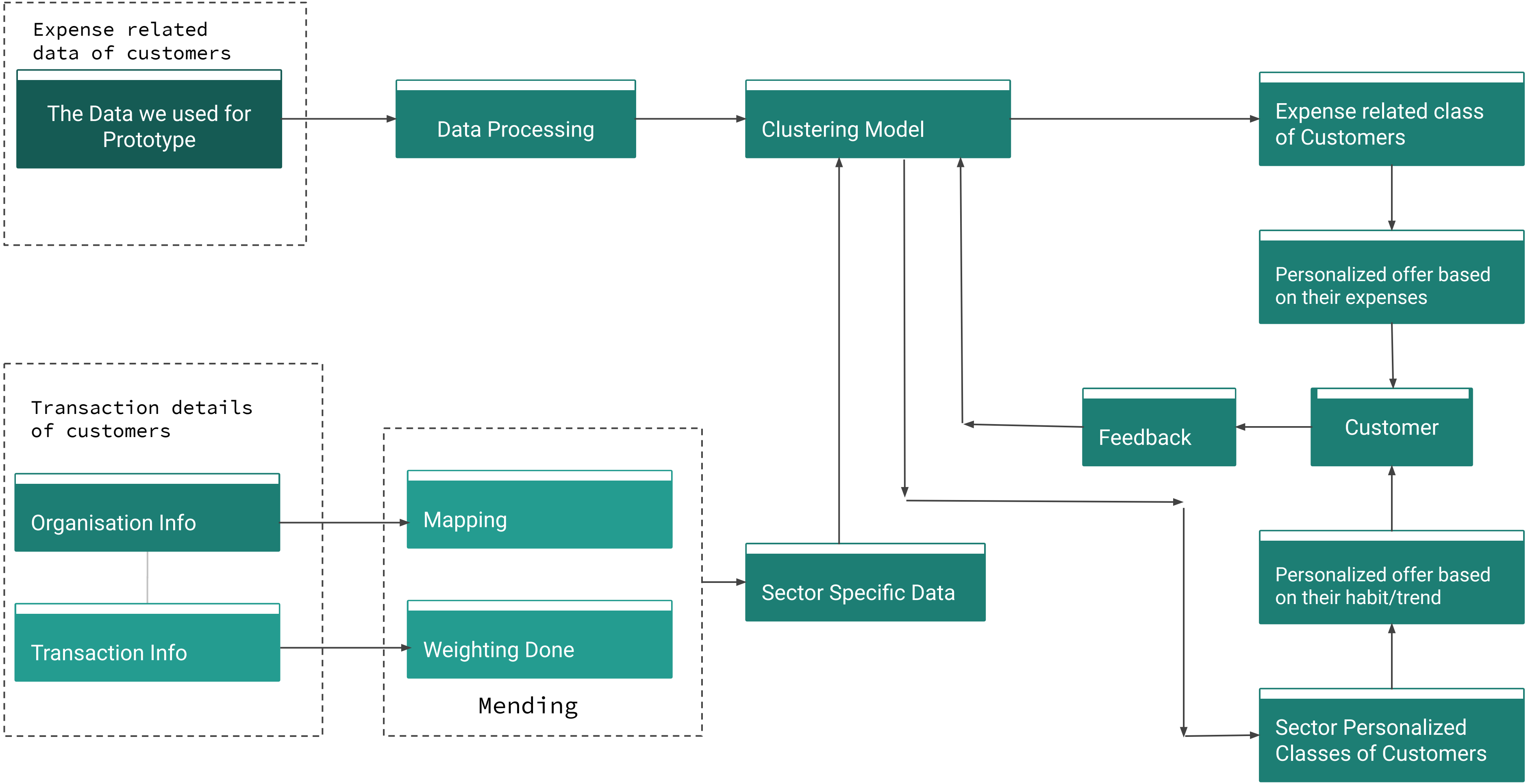
Step III: Shortlisting

- From the inferences drawn from **analysis**, best offers would be **searched**.
- These offers would be shortlisted such that they reflect the spending **habits and interests** of consumers.

Step IV: Feedback and tuning

- As new data is acquired, customer specific recommendations would be **fine tuned**.
- Also the **feedback** received from end users will be valuable for experience **enhancements**.

Let's get into Details


Expense related
data of customers
The Data we used for
Prototype
Data Processing
Clustering Model
Expense related class
of Customers
Personalized offer based
on their expenses
Transaction details
of customers
Organisation Info
Transaction Info
Mapping
Weighting Done
Mending
Sector Specific Data
Feedback
Customer
Personalized offer based
on their habit/trend
Sector Personalized
Classes of Customers

Expense-Related Offer Recommendations:

Every consumer is different on how they use credit card for their expenses. Some might not use them often while for others it might be the prominent mode of payment. So, every offer may not interest every consumer.

Our approach takes into consideration of the issue and tend to group the consumers of similar characteristics using clustering algorithm **based on their level of engagement with company's credit card as a mode of payment.** Then sending offers relatively to each group brings relevancy and personalisation into mails.

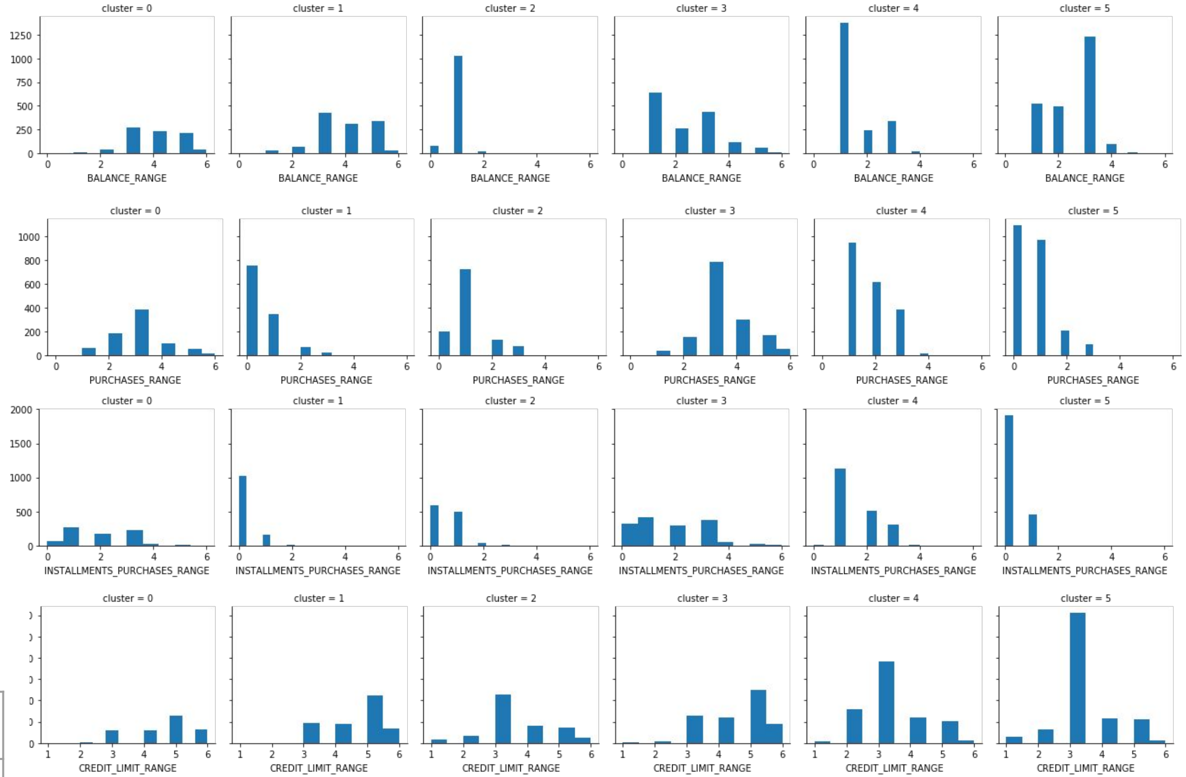
For example-> someone who has become new consumer and doesn't use company's credit card much, will tend to receive introductory offers and consumers who already use it as prime mode of payment may receive long-term plans/offers mails.

Pre-processing and Clustering

Steps Involved

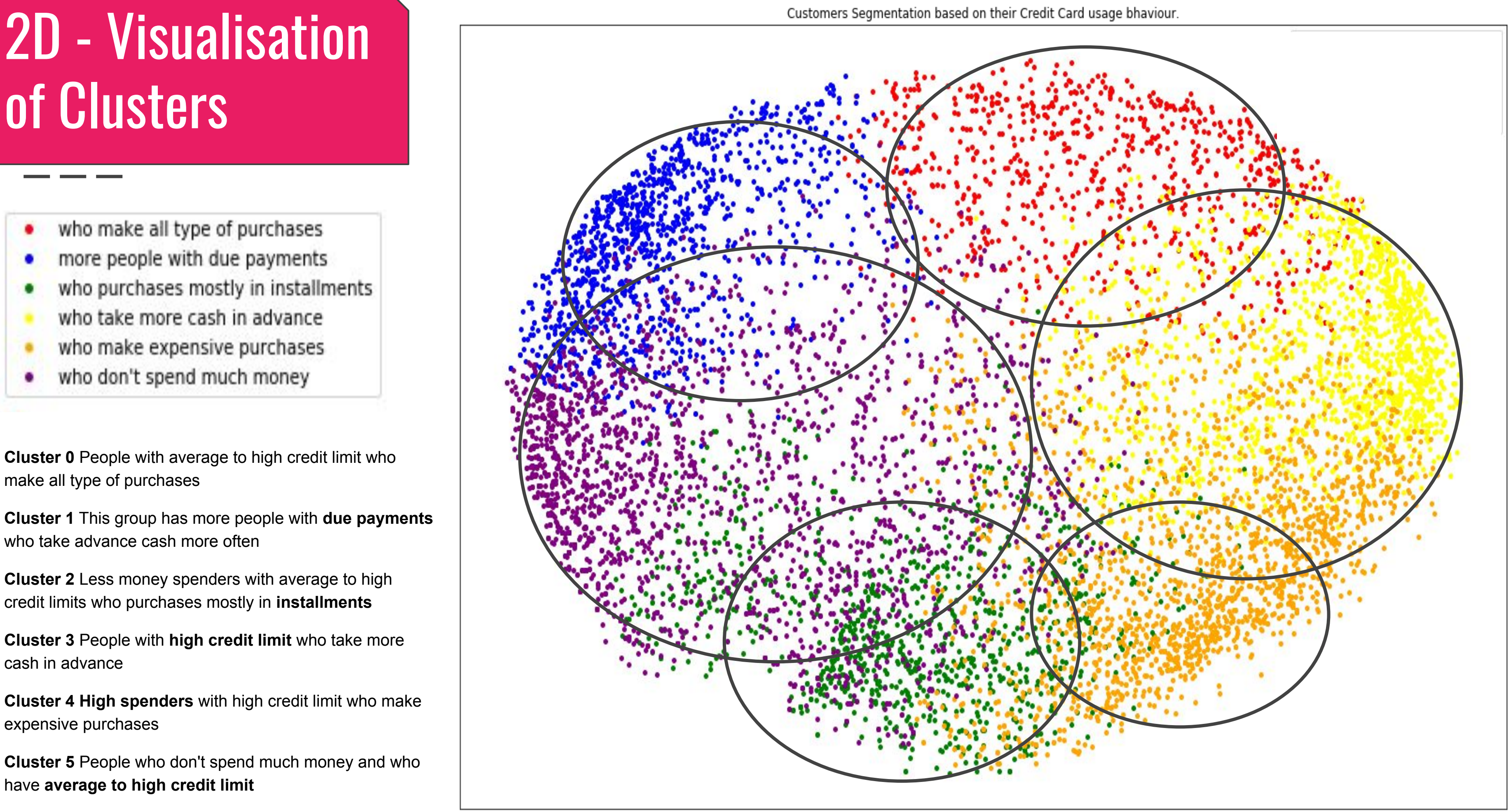
- Preprocessing of the data to deal with missing info and outliers.
- Bin the data into different ranges' bins and label them appropriately to feed them to model.
- Inputting our processed data as feature columns to the clustering algorithm.

Dataset used for prototype	[csv file]()
Algorithm used	K-Means clustering
Sample Notebook	[colab notebook]()



Some Example Graphs

2D - Visualisation of Clusters



Cluster 0 People with average to high credit limit who make all type of purchases

Cluster 1 This group has more people with **due payments** who take advance cash more often

Cluster 2 Less money spenders with average to high credit limits who purchases mostly in **installments**

Cluster 3 People with **high credit limit** who take more cash in advance

Cluster 4 High spenders with high credit limit who make expensive purchases

Cluster 5 People who don't spend much money and who have **average to high credit limit**

Sector Specific Offer Recommendations:

We don’t only personalize offers based on consumers credit card’s expense related data but also take a step ahead and personalize their offers based on where they mostly spend. To achieve this we use their transaction data and **mend** it such that consumers receive **offers personalised on their spending habit/trend.**

Here the problem is that we don't get the required data directly and even **if** we do get the data, this data cannot be fed to any model directly. Hence, here the processing or say, **mending** of data is the prime job. Thereafter, similar approach of clustering as of expense-related personalisation can be followed.

For example- if sectors predicted are food and pharmacy, and company has links to swiggy, pharmeasy, m2g then interesting and tempting offers related to these organisation when transacted via company’s credit card, can be sent.

That “MENDING” of Data

- We have transaction data of the consumers, a history of their transactions to which organisations they paid like if they bought food on zomato or did shopping on amazon, then the organization becomes zomato or amazon. The transactions which are personal or not bound to any known organisation can be listed as miscellaneous.
- Since the organizations are known, a simple mapping of the organizations to their respective sectors or their affiliations will give us sector specific transaction data on the consumers.
- Hence, a separate data-table is automanaged where the columns represent different sectors like-
    - Govt : electricity or water bills paid to govt, taxations etc.
    - Medical: transactions on pharmacy stores, medical shops or medicines bought online.
    - Food: eg- zomato, swiggy, uber-eats, restaurants etc.
    - Similarly more sectors including grocery, e-commerce, schooling, insurances-policies etc.

  Rows represent individual consumers and input is done by passing transaction data through organisation-to-sector mapping.
- Weighted inputs are done to the data-table, which weights the sector information based on the transaction amount. Then, Normalisation is done on the weighted inputs so as to handle bias/variance. *Note- normalisation parameters can be changed yearly or a couple of years later calculated on new data.*
- Thus, every cell in our data-table not only gives us individual’s sector related information but is also weighted by transaction amount. This prepared data-table, now, is fed to our unsupervised model.
- After this, same process as of expense-related-recommendations can be followed, as already shown in the above slides.

Why to use sector specific information and not organisation level data ? (i.e. why mending)

- If we do so, every organisation would then represent separate columns in data-table, which creates problem of too many data features and introduce bias in model. Also, addition/removal of any organisation in the market means new data-table and new code for the model.
    - Our solution will handle this without a bit of problem, because then we just have to state the new organisation's sector in the Mapping and done. No change in data-table. Also no bias since we don't take the organisation level info into consideration.
- let me explain with an easy example. Suppose a customer has a regular trend in transacting for food orders via zomato (for no such obvious reasons). And, our credit card company has links to Swiggy, thus, some interesting offers. Now, our model grasp the trend of consumer and sends him "food" sector offers, which then consists of Swiggy offers and the customer will now tend to use them and will use swiggy to order and the company's credit card to transact.
    - Here organisation level data won't help in any manner since the company don't have any offer of zomato.
- Keeping organisation level data may not be very acceptable by every organisation out there and unlisting them is just taking a back step.

FAQ

Is there any privacy issues?

NO, because there is no human intervention at individual consumer level data and every step is automated to the very end. Also, in case of sector-specific recommendations we map the data to sectors and do not track any individual organisation. Moreover, every data that we proposed to use are actually stored by company for use (we checked!).

Will the quality of results decrease over time?

NO, on the contrary, it will improve over time. The offers sent with our predictions will produce feedback and this feedbacks can be used to fine tune our model.

Does it require human intervention regularly?

NO, our approach is totally unsupervised. Every step is automated and once set up, no human intervention is required. Any change is very easy to incorporate.

Distribution of OFFERS

Till now, we have successfully grouped our consumers into different clusters for both categories of offers- 'expense-related' and 'sector-related'.
Now, we just need to deliver the offers in our system to right group of consumers.

- **Sector specific offers** can just be picked using sector-tags put on them. For eg- if there's a zomato offer then it would be tagged as 'Food' and kept in the system.
- Then deliver those offers to a customer which are tagged with sectors predicted for that customer. For eg- if top 3 sectors predicted for a consumer are 'pharmacy', 'food' and 'travel' then we deliver them sector-offers tagged as either of them.
- **Expense-related offers** can be picked by following 'description' of each cluster and deliver suitable ones respective to the clusters.
- 'Description' stated above is nothing but information that can be retrieved easily by analysing graphs, feature-ranges and trend of the clusters.
- Even for expense-related offers, we can do the mapping using tags if the offers are tagged while while putting in the system.
- Also, it can be followed by filtering offers with respect to obvious limitations like income and budget.

How it gets better ?

~ We call this "FEEDBACK"

We introduce our own developed mathematical SCORING function to calculate a **weighted dynamic score** which can be fed to our model as an additional feature in order to reward/penalise the predictions and thus, tune it over time. The scoring is totally dynamic and will always tend to follow consumer's trend.

Scoring function and Further details:

$$Score = \sum_{i=1}^{M} sec\_w_i \cdot sec\_rat_i + \sum_{i=1}^{N} exp\_w_i \cdot exp\_rat_i$$

Where,
M = total number of sector clusters
N = total number of expense clusters
sec_w $_i$ is weight for ith sector cluster
exp_w $_i$ is weight for ith expense cluster
sec_rat $_i$ is score-ratio for ith sector cluster
exp_rat $_i$ is score-ratio for ith expense cluster

The above stated score-ratio for ith instance is:

$$\frac{number\ of\ \ deals\ acknowledged}{number\ of\ \ deals\ provided\ +\ \in}$$

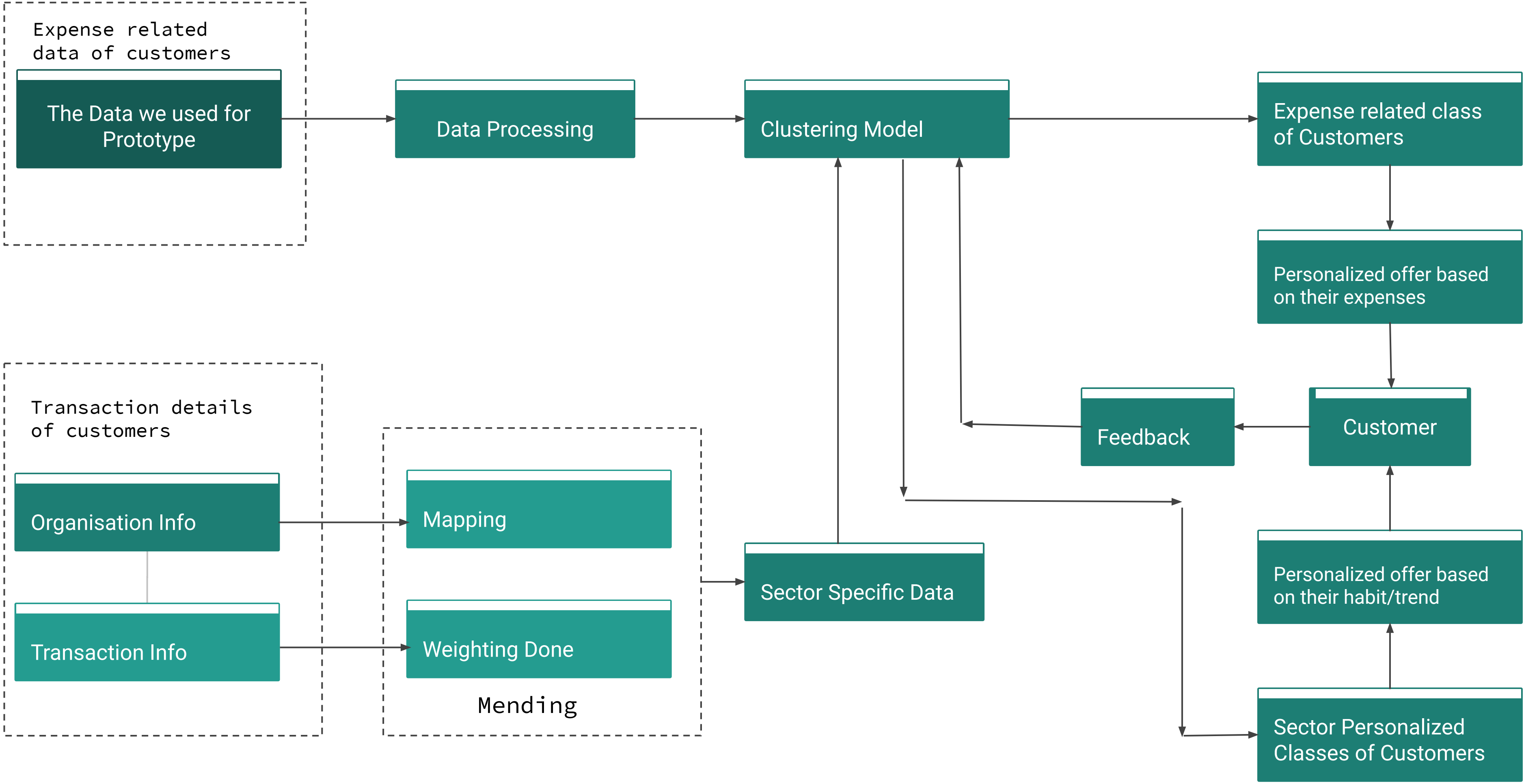
(where, ε =10^{-6} or similar to avoid division by zero, and, the offers considered are the only ones enlisted in that i^{th} instance)

- **weight** of clusters is set up in descending order such that top predictions get highest weights and lower ones get lower weights, gradually.
- If prediction diverges from optimality, the **score-ratio** gets reduced and thus final score.
- The final **score** thus calculated is fed as an additional feature to our model.
- Model learns to maximise the **score-ratio** and ultimately the final **score**.
- Taking product of both **score-ratio** and **weights** in the calculation of final score helps as it take accounts of both "**real-life-feedback**" and "**AI-predictions**".
- The score-ratio part can also be thought of as **short-term** trend and weights as **long-term** trend in consumer's activity.

What about NEW Customers?

Our database won't have the required information required by our algorithm for the new-comers.
So, How to deal with them optimally?

- We implement **Nearest-Neighbour** algorithm to find the 'k' closest data-examples available in our record with respect to new-comer.
- The data-features considered by the NN algo here, are the basic profile info (like income, profession etc) that the company take in while registration of its customers.
- Similar data-features of already enrolled consumers are sampled in the our data-space from wherein some k nearest data-points are picked using any distance norm (like euclidian).
- Then for these picked data-points, we retrieve those info which is required by our Main Model to work on.
- We calculate mean of these k data-points' info and finally represent the new-comer by this "set of info", until we collect considerable enough info on them.


Expense related data of customers
The Data we used for Prototype
Data Processing
Clustering Model
Expense related class of Customers
Personalized offer based on their expenses
Customer
Feedback
Transaction details of customers
Organisation Info
Transaction Info
Mapping
Weighting Done
Mending
Sector Specific Data
Personalized offer based on their habit/trend
Sector Personalized Classes of Customers

Our Impact

Mission: To offer bring every possible **relevancy** and **personalisation** into those simple mails.

Targeted and Customized offers

Sending personalised offers to targeted audience

Increase in Engagement

Relevant ads to customer will increase engagement

Gets better every time

Our AI model will learn and improvise using feedbacks.

The Team

Pawan Kumar Sahu
DL Research Intern | India

Pawan Raj
Android Developer | India

Pratik Bedre
DL Research Intern | India


RAKATHON
Disrupt to Dominate
Thank you