# User-Based Scheme Recommendation

# Document the Approach

## Methodology

* Custom approach using User-Based Collaborative Filtering with K-Nearest Neighbors (KNN).
* Used cosine similarity to identify similar users based on their engagement with different schemes.
* Computed an Engagement Score: log(Sales\_Value\_Last\_Period) \* (Feedback\_Score + Growth\_Percentage).
* Constructed a user-scheme interaction matrix and used sparse matrix representation for scalability.
* Generated recommendations based on the nearest neighbor's top schemes.

## Key Findings and Business Insights

* Partner behavior and engagement can effectively drive scheme recommendations.
* Highly engaged partners tend to prefer specific schemes — which can be reused to target similar profiles.
* Cosine similarity provides an interpretable metric to explain why certain schemes are recommended.
* The Engagement Score enriches recommendation logic by combining sales, feedback, and growth.

## Assumptions, Limitations, and Edge Cases

* Only one similar user is used for recommendations — can be sensitive to outlier behaviors.
* Assumes engagement score is a good proxy for interest — actual scheme redemption not validated.
* Cold-start issue exists for new users not present in the training set.
* Feedback and growth are assumed to contribute equally in the engagement formula.
* Does not account for scheme constraints like eligibility or regional availability.
* The model assumes that past scheme usage is a reliable indicator of future preference.

## Feature Descriptions

This section outlines the features used in the dataset along with their descriptions:

|  |  |
| --- | --- |
| Feature | Description |
| Partner\_id | Unique identifier for each channel partner or stockist. |
| Product\_id | Product associated with the scheme engagement. |
| Geography | Region or location of the stockist. |
| Stockist\_Type | Type/category of the stockist (e.g., distributor, wholesaler). |
| Scheme\_Type | The type/category of scheme used. |
| Sales\_Value\_Last\_Period | Sales value achieved by the partner in the last time period. |
| Sales\_Quantity\_Last\_Period | Quantity of items sold by the partner in the last time period. |
| MRP | Maximum Retail Price of the product. |
| Growth\_Percentage | Percentage growth in sales compared to a previous baseline period. |
| Feedback\_Score | Numerical rating given by the partner indicating satisfaction or product feedback. |
| Engagement\_Score | Derived feature representing partner engagement based on sales, feedback, and growth. |

## References

The following references and documentation resources were helpful while implementing the approach:

* **User-Based Collaborative Filtering - Surprise Library:** http://surpriselib.com/
* **KNN for Recommendation Systems - scikit-learn:** https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html
* **Pandas DataFrame Pivot Table:** https://pandas.pydata.org/docs/reference/api/pandas.pivot\_table.html
* **Scipy Sparse Matrices:** https://docs.scipy.org/doc/scipy/reference/sparse.html
* **Random Forest Feature Importance:** https://scikit-learn.org/stable/auto\_examples/ensemble/plot\_forest\_importances.html