**ZEPTO – PS**

**Project Documentation: CTR Prediction and Product Ranking**

1. **Project Overview** :

This project aims to predict the **Click-Through Rate (CTR)** and rank products based on search terms and city levels, utilizing classical machine learning techniques. The objective is to enhance user experience and product visibility on the platform.

2**. Objective**

- CTR Prediction: Estimate the likelihood of a product being clicked by a user.

- Product Ranking: Rank products according to **relevance** to search terms and city-specific preferences.

3. **Data Description**

- Data of the platform, including user interactions and product details.

- Features:

- Numeric: latest margin , savings , total clicks etc.

- Categorical: Query type, product category, product name , product subname etc.

4. **Data Preprocessing**

1. Exploratory Data Analysis (EDA):

- Analyzed data distributions, relationships, and trends.

- Identified and addressed missing and null values.

2. Handling Missing Values:

- Numeric Data: No null values were found.

- Categorical Data: Null values were handled by encoding the **'query\_type'** feature. Other less significant categorical features were removed to reduce complexity and computational load.

3. Feature Selection:

- A correlation matrix was utilized to identify the relationships between features.

- Features with an absolute correlation value with the target variable **('is\_clicked')** less than **0.05** were removed, as they contributed minimally to the prediction.

4. Standardization:

- Important features were standardized using a standard scaler, applying the **z-score** method. This ensured that feature values were comparable, aiding in achieving more consistent model performance.

5**. Modeling**

- Train test split was done with test\_size=0.25 and model was fitted with train data.

- Algorithm Used:

- **Logistic Regression** model was used for the classification task .

- Hyperparameter Tuning was done with regularization to get the best parameters. Regulariztion strength -**100** and **l2** regularization were the best parameters selected.

- Evaluation Metrics:

- Model performance was measured using accuracy and Area Under the Curve (AUC) metrics. Achieved accuracy - **0.92** and auc score - **0.88**

6**. Product Ranking Methodology**

1. Classification and Probability Estimation :

- Products were classified based on search terms and city\_id.

- **Predicted probabilities** for the target variable ('is\_clicked') were calculated.

2. Ranking Criteria:

- Products were ranked using a **pointwise ranking** method, based on predicted probability and query\_product\_similarity.

- If products had the same predicted probability, query\_product\_similarity was used as a tiebreaker.

3. Evaluation Metric :

- The quality of the ranking was assessed using the **Mean Average Precision (MAP)** score, which evaluates the precision of the top-ranked items.

7. **Results and Insights** :

-The ranking system effectively differentiated products based on relevance, providing a meaningful and user-centric order.