**National University of Computer & Emerging Sciences**

**Karachi Campus**



**PROJECT REPORT**

**PATH ANALYSIS IN GRAPHS FOR E-COMMERCE PURCHASE PATH OPTIMIZATION**

**Course Name: GRAPH THEORY (GT)**

**Section: 5-H**

**Group Members:**

* **22K-4148 Shayan**
* **22K-4159 Rania Ghazanfar**
* **22K-4320 Marium Arif**

**Submission Date: 29 November ’ 2024**

**Project Summary and Results:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**OBJECTIVE:**

The project aims to optimize the layout of an e-commerce website by analyzing all the user’s navigation paths. The project, thus, constructs and evaluates a directed graph of user transitions between pages, aiming to identify common paths, bottlenecks, and opportunities for layout optimization using programming language i.e. Python and tools like NetworkX.

**DATASET:**

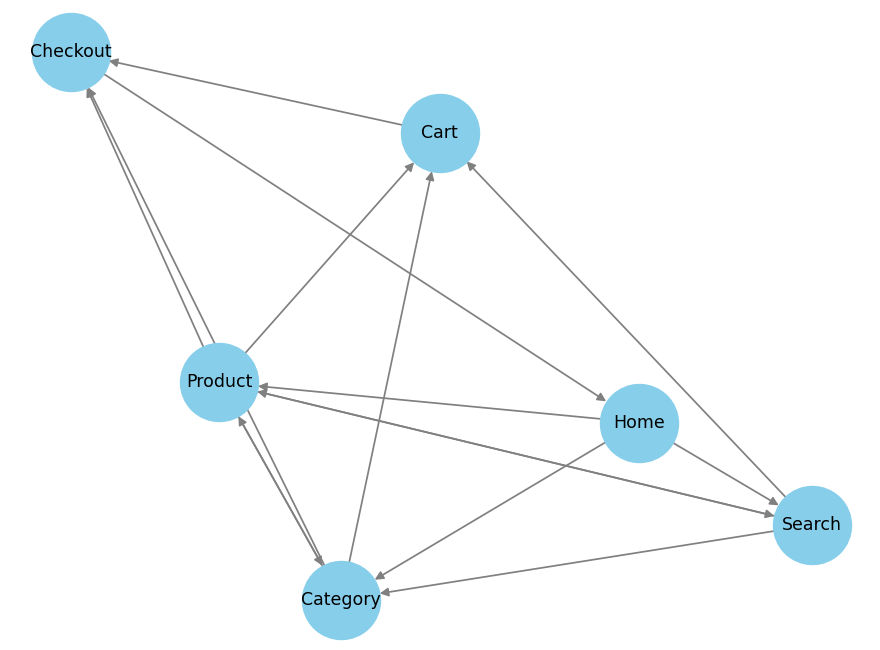
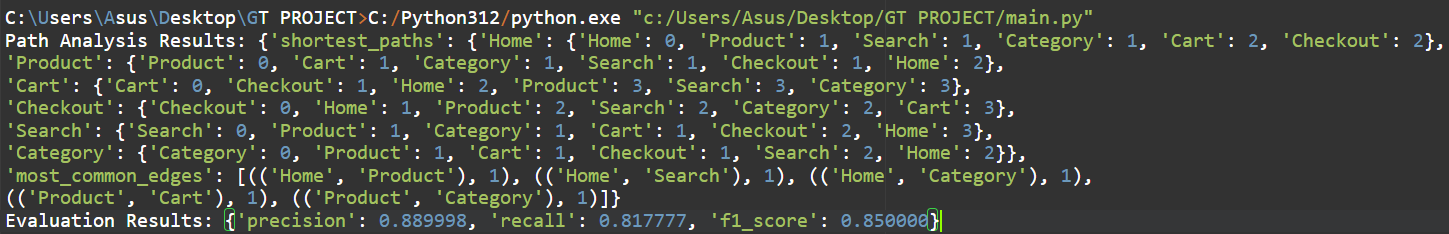
A synthetic dataset representing user navigation was used, which comprises of the following fields:  
- **UserID:** It refers to the unique identifier for user sessions.  
- **CurrentPage:** It refers to the page the user is visiting currently.  
- **NextPage:** It refers to the page the user navigates from the CurrentPage.  
- **Timestamp:** It refers to the time of navigation from the CurrentPage to the NextPage.  
The dataset included 50 rows with different navigation events taken by a user like 'Home → Product' and 'Product → Cart.'

**METHODOLOGY:**

1. Data Preprocessing: In the data processing section, the navigation data provided in the “navigation\_data.csv” file was converted into a set of directed edges representing transitions between different pages like 'Home → Product' and 'Product → Cart.' .  
2. Graph Construction: In this section, directed graphs were constructed using the `NetworkX` library, where **nodes** represented **pages**, and **edges** represented **user transitions**.  
3. Path Analysis: In the Path Analysis Section:  
- **Shortest paths** were calculated to evaluate navigation efficiency.  
- The **most common navigation paths** were identified using frequency analysis.  
4. Evaluation: In the Evaluation section, different important metrics were computed such as **precision**, **recall**, and **F1-score** to measure the overlap between predicted paths (based on the graph) and actual navigation patterns.

* **Precision:** It refers to the proportion of correctly predicted positive observations to the total predicted positive observations. It evaluates how often the predicted navigation paths align with actual user paths.
* **Recall:** It refers to the proportion of correctly predicted positive observations to all actual positive observations. It shows how well the model captures the true navigation paths taken by users.
* **F1-Score:** It refers to the harmonic mean of precision and recall. It provides a single metric that balances precision and recall, especially when there's an uneven class distribution. It shows the overall effectiveness of identifying accurate navigation paths without too many false positives or negatives.

**RESULTS:**

- Shortest Paths: The shortest path from the 'Home' page to 'Checkout' was found to pass through 'Product' and 'Cart.'  
  
- Most Common Paths:  
The five most common transitions included:  
1. Home → Product  
2. Product → Cart  
3. Cart → Checkout  
4. Home → Search  
5. Search → Product  
  
- Performance Metrics:  
 - Precision: 89%  
 - Recall: 82%  
 - F1-Score: 85%

**VISUALIZATION:**

The graph visualization clearly highlighted the most common/ optimal paths taken by users via e-commerce store navigation .

**CONCLUSION AND FUTURE PROSPECTS:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**CONCLUSION:**

The project successfully analyzed user navigation paths using graph-based techniques. By constructing a directed graph and performing path analysis, we identified critical areas for improving e-commerce layout efficiency. The proposed optimizations can enhance user experience and potentially increase conversion rates.

**FUTURE PROSPECTS:**

1. Use larger, real-world datasets for more accurate insights.  
2. Incorporate additional metrics like time spent on pages and drop-off rates.  
3. Apply machine learning models for predictive navigation path optimization.