DEPARTMENT OF APEX INSTITUTE OF TECHNOLOGY

PROJECTPROPOSAL

1. Project Title: - Segmentation study on Bank customers based on RNN.

2. Project Scope: - Customer segmentation is a vital strategy for businesses to better understand and target their diverse customer base. While traditional segmentation methods are effective, they might not fully capture the temporal aspects of customer behavior. This project proposes the utilization of Recurrent Neural Networks (RNNs) to perform customer segmentation, focusing on sequential data analysis.

Scope of the project:

1. **Data Collection and Preprocessing:**

   - Gather the customer data you need for segmentation. This could include attributes like purchase history, demographic information, browsing behavior, etc.

   - Preprocess the data to handle missing values, outliers, and normalize/standardize numerical features.

2. **Traditional Clustering Methods**:

   - Implement K-means clustering to divide customers into groups based on their attributes.

   - Use of density-based clustering algorithms like DBSCAN to identify clusters of varying shapes and densities.

   - Implement model-based clustering techniques such as Gaussian Mixture Models (GMM) to find clusters that follow probabilistic distributions.

3. **Recurrent Neural Network (RNN)**:

   - Convert the customer data into sequences if it's sequential in nature (e.g., time-series data).

   - Design an RNN architecture suitable for your data. You may use libraries like TensorFlow or PyTorch in R to build and train your RNN model.

   - Train the RNN on the customer data. The RNN might capture temporal patterns or sequential dependencies in the data.

4. Combining Clustering with RNN:

- Once you have the results from traditional clustering methods and the RNN, consider how you want to combine these results. You can use the clusters from traditional methods as additional features for the RNN or vice versa.

- Ensemble techniques, like averaging or stacking, can be employed to combine the results from different methods.

5. Evaluation:

- Assess the quality of your segmentation using appropriate metrics (e.g., silhouette score, Davies-Bouldin index, etc. for traditional clustering; appropriate evaluation metrics for RNN).

- Cross-validation can help ensure the stability and robustness of your results.

3. Requirements: -

- Hardware Requirements
  - **GPU Acceleration**
  - **Server:**
    - Processor: Quad-core or higher, modern processors (e.g., Intel Core i5 or equivalent(i3 11th gen)
    - Memory (RAM): 8GB or more
    - Storage: 256GB SSD or higher
  - **Network**
  - **Database:**
    - Storage: 500GB HDD or SSD
    - Database Software: MySQL or PostgreSQL
  - **Monitoring and Management**

- Software Requirements

  - 1.**Operating System**(Mac OS, Windows OS)
  - 2. **Development Environment:**
    - Integrated Development Environment (IDE) like Visual Studio Code, PyCharm, or Jupyter Notebook for software development and testing.

  - **Clustering Libraries**:

  - **R Packages**: You'll need various R packages for traditional clustering methods, such as `stats` for basic statistics and `cluster` for clustering algorithms.

  - **Python Libraries**: For clustering in Python, you might use libraries like `scikit-learn` for K-means and DBSCAN, and `mixtools` for model-based clustering.

Deep Learning Frameworks:

- **TensorFlow**: For building and training neural networks, including RNNs.

PyTorch: An alternative to TensorFlow, especially popular for its flexibility in   research settings.

- **Keras**: A high-level neural networks API that runs on top of TensorFlow or Theano, offering a user-friendly interface.

STUDENTS DETAILS

Name	UID	Signature
Ishpreet Kaur	21BCS6424	
Vaibhav Jaitwal	21BCS6454	

APPROVAL AND AUTHORITY TO PROCEED

We approve the project as described above, and authorize the team to proceed.

Name	Title	Signature (With Date)
Dr. Priyanka Kaushik	Supervisor	