**Week 6**

**METHODOLOGY**

The development of the proposed real estate system involves multiple stages, including data collection, preprocessing, feature engineering, model development, and front-end integration. City-specific datasets were first cleaned and processed to remove inconsistencies, missing values, and outliers. For the recommendation system, TF-IDF (Term Frequency-Inverse Document Frequency) vectorization was applied on textual data such as location and facility descriptions. Cosine similarity was then used to measure the closeness between property listings and the user's preferences, allowing the system to generate personalized property recommendations.

For price prediction, separate Random Forest Regression models were trained for each city using key numerical features such as total square footage, number of bathrooms, balconies, or BHK depending on the city. These models were evaluated using standard metrics like RMSE and then serialized using joblib for deployment. The front-end of the system was developed using Streamlit, which provides an interactive interface where users can select a city, explore recommended properties, and input property features to get real-time price predictions. All components were integrated seamlessly, ensuring a smooth and efficient user experience while maintaining city-specific accuracy in both recommendations and predictions. To enhance modularity and maintainability, the system architecture was designed with a clear separation between the backend logic and the user interface. Each city’s data and models were handled independently to allow city-specific tuning and future scalability. The recommendation module dynamically loads the appropriate dataset and similarity matrix based on the selected city, while the prediction module loads the corresponding trained Random Forest model. Streamlit widgets such as dropdowns, sliders, and input fields were utilized to gather user inputs, which are then passed to the backend functions for generating outputs. The system ensures responsiveness by performing all computations on-the-fly and presenting results in a visually intuitive format, making it suitable for both technical and non-technical users exploring real estate options.