**Week 3**

**LITERATURE SURVEY**

The use of machine learning in real estate analytics has gained significant traction in recent years. Researchers and developers have explored various models and techniques to predict property prices and recommend suitable listings to users. One of the earliest approaches involved Linear Regression, where the relationship between price and features like area, number of rooms, and location was modeled linearly. However, this model struggled with non-linear data patterns and outliers. Later, ensemble methods such as Random Forest and Gradient Boosting Machines became popular due to their ability to model complex relationships and deliver higher accuracy. In terms of recommendation systems, two primary categories are observed: collaborative filtering and content-based filtering. While collaborative filtering uses user interaction history, it is limited in real estate due to sparse and unique property listings. Content-based filtering, on the other hand, uses property attributes and user preferences to generate recommendations, making it more suitable for this domain.

**Review of Existing Systems**

**1. 99acres, MagicBricks, and Housing.com**

These commercial real estate platforms provide property listings, filters based on user input, and basic price trends. However, they often lack advanced price prediction features and personalized recommendations. They do not utilize machine learning models for price forecasting or similarity-based suggestions in a transparent or customizable manner.

**2. Kaggle-Based Projects**

Multiple open-source projects on Kaggle focus on house price prediction using regression techniques like Random Forest, XGBoost, or LightGBM. These projects often demonstrate good accuracy but are typically limited to a single city (commonly Bangalore or Boston datasets) and are not integrated into a user-friendly web interface.

**3. Academic Research Prototypes**

Some academic models propose advanced neural networks or hybrid recommendation systems combining NLP and user behavior data. While accurate, these systems are often not production-ready due to complexity and lack of frontend integration.

**Limitations in Existing Systems**

* Lack of city-wise modular implementation.
* Absence of real-time recommendation based on dynamic user input.
* No integration of both recommendation and price prediction in a single platform.
* Minimal use of intuitive web technologies like Streamlit for deployment.

Our proposed system overcomes these limitations by:

* Providing city-specific recommendations and predictions.
* Using TF-IDF + Cosine Similarity for personalized suggestions.
* Deploying a full system via Streamlit, allowing for real-time interaction.
* Offering a scalable and modular design that can easily integrate more cities or features.