Smarter Property Investments

The GenG Model for Property Investors (GGMPI)

Presented by: Alham Hotaki

Date: 09.09.2024

Making Informed Decisions in an Unpredictable Market

- **Challenge for Investors: Real estate is a complex and unpredictable market where investing in the wrong property can lead to significant financial losses.**
- **Volatility, hidden costs, and long-term risks make it difficult to assess the potential of a property investment.**

Key Question:

How can you consistently make smarter, data-driven property investment decisions?

Introducing the GenG Model for Property Investors (GGMPI)

Your Personalized Risk Analysis Tool

- What if you had a model that provides a clear, quantitative risk assessment for any property?
- The GenG Model does exactly that by:
  - Quantifying risk based on key financial and property-related factors.
  - Classifying properties into High, Medium, or Low risk.
  - Empowering you to make informed investment decisions backed by data.

Key Risk Factors Investors Should Know

Price Volatility

How much have property prices fluctuated in the area? High volatility means higher risk.

Price per SQFT

Expensive properties may seem appealing, but they carry a higher risk due to larger financial commitments.

Maintenance Costs

Taxes and HOA fees impact long-term profitability.

Age of Property

Older properties can incur unexpected repair and maintenance costs.

Your Competitive Edge with GenG Model

- GGMPI offers more than just data; it provides a clear risk profile for each property, allowing you to:
  - Avoid high-risk properties.
  - Focus on properties that offer stability and long-term growth.
  - Save time and effort by focusing your investments in the right areas.
- Make smarter investment choices with a model designed to minimize risks and maximize your returns.

How GenG Model works?

Risk Profile Definition

High Risk

- Top 25% in sold price.
- High price volatility.
- Older properties (built before 1980).
- High maintenance costs.

Medium Risk

- Middle 50% in sold price, volatility, and maintenance costs.
- Built between 1980 and 2000.

Low Risk

- Bottom 25% in sold price, volatility, and maintenance costs.
- Newer properties (built after 2000).

Weighting System for Risk Calculation

To calculate the final risk score, we assigned weights to the factors based on their impact:

- **Price Volatility:** 50%
- **Sold Price:** 30%
- **Maintenance Cost:** 10%
- **Property Age:** 10%

This formula provides a **weighted score** which is then mapped to the **High**, **Medium**, or **Low** risk categories based on thresholds.

Feature Engineering

To improve prediction accuracy, several features were engineered:

- **Price Per Square Foot:** Normalizes the sold price relative to the size of the property.
- **Price Volatility (CV):** Calculated as the coefficient of variation within each ZIP code. It's crucial for identifying price fluctuations.
- **Maintenance Cost:** Sum of property taxes and HOA fees, impacting the ongoing cost of ownership.

Risk Classification Workflow

The model classifies properties into risk categories using the following steps:

1. Calculate each property's **weighted score** based on volatility, sold price, maintenance cost, and age.
2. Classify the property based on the weighted score:
   - **High Risk**: Weighted score ≥ 2.5
   - **Medium Risk**: Weighted score between 1.5 and 2.5
   - **Low Risk**: Weighted score ≤ 1.5

Model Implementation with KNN

The **K-Nearest Neighbors (KNN) model** was chosen to predict risk based on similar properties.

- **Training:** The model was trained on a dataset using features like price per square foot, property age, maintenance costs, and price volatility.
- **Prediction:** The KNN model finds the closest properties based on these features and predicts the risk level for a new property.
- **Accuracy:** The model achieves around **80% accuracy** in predicting investment risk on test data.

Geospatial Calculation for Missing Locations

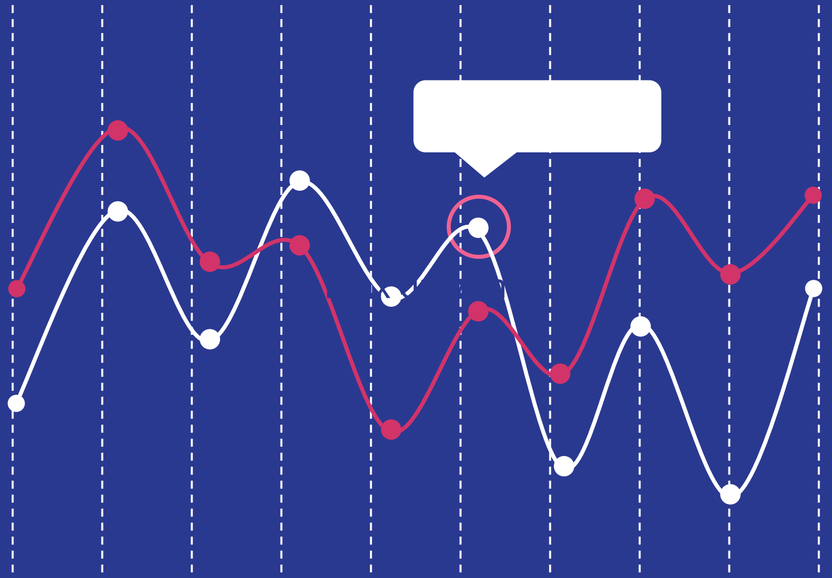
In cases where property coordinates (latitude and longitude) are missing from the dataset:

1. The model calculates price volatility by finding nearby properties using **Haversine distance** (geographical distance).
2. It gradually expands the search radius to find the nearest properties.
3. This method ensures that even if exact data isn't available, the model can still make an informed prediction.

Expertise of the GenG Model

1. Provides quantitative, data-driven risk assessments for real estate investments.
2. Incorporates economic factors, property-specific features, and market volatility into risk prediction.
3. The use of KNN and geospatial analysis ensures reliable predictions, even with incomplete data.
4. GGMPI is a powerful tool for investors to assess potential risks and make informed decisions in the property market.

Thank You



What is K-Nearest Neighbors (KNN)?

What is K-Nearest Neighbors (KNN)?

K-Nearest Neighbors (KNN) is a **supervised machine learning algorithm** used for classification and regression tasks. It's based on the idea that similar data points are close to each other.

Why Use KNN?

1. **Simplicity**: It's easy to understand and implement.
2. **Versatility**: It can be used for both classification (categorizing data) and regression (predicting continuous values).
3. **Non-parametric**: It doesn't make any assumptions about the underlying data distribution.

How Does KNN Work?

1. **Choose the number of neighbors (K)**: This is the number of closest data points the algorithm will consider.
2. **Calculate the distance**: Measure the distance between the new data point and all existing data points using a distance metric (e.g., Euclidean distance).
3. **Find the K nearest neighbors**: Identify the K data points that are closest to the new data point.
4. **Make a prediction**: For classification, the new data point is assigned to the class that is most common among its K nearest neighbors. For regression, the prediction is the average of the values of its K nearest neighbors.

Example

Imagine you have a dataset of people's heights and weights, and you want to predict their T-shirt size (Small, Medium, Large).

Table

Height (cm)	Weight (kg)	T-shirt Size
158	58	M
160	59	M
163	60	M
165	65	L
170	68	L

Now, you have a new person with a height of 161 cm and weight of 61 kg, and you want to predict their T-shirt size.

1. **Choose K**: Let's say K = 3.
2. **Calculate distances**: Compute the distance between the new person and all existing data points.
3. **Find the 3 nearest neighbors**: Suppose the 3 closest people are:
   - (160, 59) with size M
   - (163, 60) with size M
   - (165, 65) with size L
4. **Make a prediction**: The majority of the nearest neighbors (2 out of 3) have size M, so the new person is predicted to have size M.