Robou: A Land Market Intelligence and District Recommendation Platform

Sarah Dhaifallah AlOtaibi
Imam Abdulrahman Bin Faisal University
College of Computer Science and Information Technology

Hoton Aoun AlQahtani
Imam Abdulrahman Bin Faisal University
College of Computer Science and Information Technology

Reema Ali AlSahli
Imam Abdulrahman Bin Faisal University
College of Computer Science and Information Technology

Layan Abdullah AlQahtani
Imam Abdulrahman Bin Faisal University
College of Computer Science and Information Technology

Fatimah Alaa AlMumtan
Imam Abdulrahman Bin Faisal University
College of Computer Science and Information Technology

Deemah Nabeel AlOmair
Imam Abdulrahman Bin Faisal University
College of Computer Science and Information Technology

Suhier Bashier Elfaki
Imam Abdulrahman Bin Faisal University
College of Computer Science and Information Technology

ABSTRACT
Selecting suitable land can be a challenging process because market information is scattered across multiple sources and presented in inconsistent formats. Many users rely on manual research, informal advice, and personal judgment when evaluating land, which makes the process time-consuming, less structured, and difficult to compare across districts. Robou is a bilingual web-based land market intelligence and district recommendation platform designed to support smarter land selection in the Eastern Province of Saudi Arabia, specifically Dammam, Khobar, and Dhahran. The platform integrates market data, geographic indicators, interactive maps, and machine learning to provide organized, location-based insights within a single system. Users can select an area directly on the map, choose land type, and define preferred proximity to services. Based on these inputs, Robou evaluates districts and returns ranked recommendations supported by district-level market data and predictive analysis. The system also supports secure sign-in and personalized features such as saving favorite districts for later review. The final machine learning model uses CatBoostRegressor, where the best performance was achieved when the minimum number of deals was set to 10. The model achieved a test R^2 of 0.645, MAE of 359.90 SR/m², and MAPE of 20.77%, showing improved ability to learn district-level price patterns and support the recommendation process.

General Terms
Machine Learning, Geographic Information Systems, Real Estate Analytics, Decision Support Systems, Recommendation Systems.

Keywords
Robou, land market intelligence, district recommendation, CatBoostRegressor, GIS, Eastern Province, Saudi Arabia, real estate analytics.

1. INTRODUCTION
Selecting suitable land is an important decision for individuals and investors, but the process is often difficult because land market information is distributed across different sources and presented in different formats. Users may need to manually compare districts, check prices, review maps, and rely on informal advice before deciding. This makes land selection time-consuming and less reliable, especially when users want to compare districts based on price trends, land purpose, and nearby services.

Robou addresses this problem by providing a bilingual web-based platform for land market intelligence and district recommendation in Saudi Arabia's Eastern Province. The platform focuses on Dammam, Khobar, and Dhahran, where users can explore land opportunities through interactive maps and structured district-level insights. Instead of presenting raw market information only, Robou combines market indicators, geographic indicators, and machine learning to help users understand price differences and identify suitable areas more clearly.

The main contribution of Robou is the integration of spatial selection, user preferences, predictive analysis, and recommendation logic in one practical system. Users can select a target area on the map, choose whether the land purpose is residential or commercial, and define preferred proximity to services. The platform then evaluates districts within the selected area and provides ranked recommendations supported by market data and predicted price guidance. This helps reduce reliance on scattered manual analysis and supports more informed land-related decisions.

2. RELATED WORK
Existing real estate and land selection tools provide useful market information, but many of them do not fully support localized district-level decision making for everyday users. General market evaluations and national housing initiatives highlight the growing demand for structured land data, yet information remains heavily fragmented across high-level institutional overviews and broader sector assessments [1, 7]. Land market information is often scattered across multiple sources, which makes it difficult for users to compare districts and understand

price trends in a structured way. This data fragmentation is further compounded by the reliance on traditional market indexes and macro-level financial reports that capture national or regional trends rather than neighborhood-specific dynamics [2, 8, 9]. Consequently, many platforms present raw listings or general property data without meaningful recommendation support, and users still need domain knowledge to interpret the information and decide which district is more suitable.

The gap is clearer in the Eastern Province of Saudi Arabia, where few localized platforms focus on district-level land market intelligence. While geographic information systems (GIS) have been introduced to analyze real estate services and map property distribution in other regions of the Kingdom, their application remains mostly restricted to administrative or mobile-based mapping without specialized decision-support layers [3, 4]. Current solutions often lack integration between user preferences, spatial selection, market indicators, and recommendation output. They also do not always combine interactive maps with machine learning-based price guidance. Although recent research has begun deploying artificial intelligence and advanced machine learning algorithms to predict property prices in Saudi Arabia, these predictive models are rarely integrated into consumer-facing, localized spatial recommendation platforms [5, 6].Therefore, there is a need for a platform that transforms fragmented data into organized, location-specific insights and provides users with ranked district recommendations based on their selected criteria.

3. PROPOSED SYSTEM

Robou is proposed as a bilingual web-based platform that supports land market intelligence and district recommendation in Dammam, Khobar, and Dhahran. The system allows users to interact with a map, select the area they are interested in, choose the land type, and define preferred proximity to services. Based on these inputs, the platform evaluates districts and returns ranked recommendations that are supported by district-level market data and predictive analysis.

The proposed solution combines four main components: interactive geographic selection, market indicators, machine learning-based price guidance, and user-centered recommendation logic. This enables users to explore districts more easily, compare areas effectively, and make more informed and confident decisions. The platform also includes secure sign-in and personalized features such as saving favorite districts for later review, creating a more consistent and user-centered experience.

3.1 System Objectives

The objectives of Robou are to develop a bilingual web platform for land market intelligence, support users in selecting suitable districts for residential or commercial purposes, integrate interactive maps with market indicators and recommendation logic, apply machine learning to improve district-level price guidance and recommendation quality, and provide secure user access with personalized features such as saved favorite districts.

These objectives are designed to reduce manual research, improve district comparison, and provide clearer support for land-related decisions. By combining user preferences, spatial information, market data, and predictive analysis, Robou aims to make land selection more structured, transparent, and practical for users in the Eastern Province.

3.2 System Scope

The scope of Robou focuses on land market intelligence and district recommendation for three cities in Saudi Arabia's Eastern Province: Dammam, Khobar, and Dhahran. The platform supports users who want to evaluate districts for residential or commercial land selection. It does not focus on land selling, listing creation, or administrative land management; instead, it focuses on helping users compare areas and identify suitable districts based on their selected map area, land type, service proximity, and district-level market indicators.

Within this scope, the system provides interactive map-based area selection, district comparison, ranked recommendations, predictive price guidance, secure authentication, and saved favorite districts. The machine learning component is used to improve price guidance at the district level. In the final experiment, CatBoostRegressor achieved the best performance with min_deals = 10, producing a test R^2 of 0.645, MAE of 359.90 SR/m², and MAPE of 20.77%. These results indicate that the model can capture district-level price patterns more effectively than the baseline and provide useful support for the recommendation process.

4. METHODOLOGY

The proposed system, Robou, was developed as a web-based land valuation and district recommendation platform for the Eastern Province of Saudi Arabia (Dammam, Khobar, and Dhahran). It helps users define an area on a map, specify preferences (residential or commercial land, budget per square meter, proximity to facilities), and receive ranked district recommendations with estimated price per square meter, growth indicators, and short explanatory reasons. The methodology covers data collection and aggregation, preprocessing and feature construction, machine learning for price estimation, multi-criteria ranking, interactive analytics, and map-based interaction.

4.1 Data Collection

The system relies on real-estate transaction data and supporting geographic datasets, prepared through an offline pipeline before deployment.

Market transaction data comes from publicly available real-estate indicators and manually organized datasets (e.g., sales spreadsheets converted to structured tables). Records include city, district, property type, transaction price, area, and date, from which price per square meter is calculated. The focus is on land parcels for residential and commercial use.

District and location data include standardized district names, city labels, and geographic coordinates (centroids) for whitelisted neighborhoods in the three target cities.

Supplementary spatial data describe access to services: distances and counts of schools, hospitals, and commercial facilities, obtained from open map sources and/or places APIs, to characterize neighborhood accessibility.

Aggregated analytics store district-level statistics by year and quarter (e.g., median price per m², number of deals, price dispersion) and year-over-year growth at district and city level.

Processed data are stored in two roles:

- A relational database (MySQL) holds districts, quarterly aggregates, growth tables, historical sales, and model metadata for the ML service.

- Supabase (PostgreSQL) supports optional analytics mirroring, user authentication, profiles, and saved favorites.

This separates market analytics from user-specific data.

4.2 Data Preprocessing

Before training and serving predictions, preprocessing improves consistency and reliability:

- Filtering: retain land-related transactions; exclude unrealistic price-per-square-meter values outside an acceptable range.
- Cleaning: remove duplicates; handle missing or invalid numeric fields.
- District normalization: standardize Arabic district names and map raw labels to a whitelist of districts with valid coordinates; exclude ambiguous or incomplete entries.
- Temporal aggregation: group transactions by city, district, property type, year, and quarter to produce stable quarterly medians and deal counts.
- Quality thresholds: require minimum transaction counts per district-quarter so sparse areas do not distort results.

Invalid or unmapped records are excluded so recommendations and charts reflect verified districts only.

4.3 Feature Engineering

Features are built at district–quarter level for machine learning and ranking:

- Price features: median price per m², baselines, and residual/log transforms for modeling.
- Market activity: number of deals, price spread (standard deviation, interquartile range), min and max prices in the quarter.
- Growth features: year-over-year change in median price per district and city.
- Location features: district centroid latitude and longitude.
- Services/accessibility: distances to nearest school, hospital, and mall; counts of facilities within a fixed radius.
- User-facing inputs: land use (residential/commercial), proximity to facilities (near, medium, far), and a reference plot area (e.g., 400 m²) for displaying total estimated value.

A gradient boosting regressor (e.g., CatBoost) is trained on these features to predict aggregated residual price per square meter, producing a serialized model loaded by the API at startup.

4.4 Recommendation Logic

The module recommends districts within a user-selected geographic area, not individual land listings from a searchable catalog.

User inputs on the predict page:

- A bounding box drawn on the map.
- Land use: residential or commercial.
- Budget: minimum and maximum price per m² (SAR).
- Proximity to facilities: near, medium, or far.
- Optional reference area for total price display.

Processing steps:

1. Geographic filtering : Select districts whose centroids lie inside the drawn area (limited to the three supported cities).
2. Price prediction : Estimate price per m² per district using the ML model on the latest available year and quarter.
3. Budget filter : Exclude districts outside the user's price range.
4. Multi-criteria scoring (0–100) , Combine:
   - Price score : affordability within the range (weight depends on mode).
   - Growth score : from year-over-year growth.
   - Services score : from facility/accessibility index, adjusted by proximity preference.
5. Ranking modes : *Value* (price-focused), *premium* (services-focused), or *growth* (growth-focused), each with different component weights.
6. Output : Top districts (typically three), each with price per m², growth trend, services level, confidence, and two short Arabic reasons (e.g., suitable price, above-average services).

Logged-in users can save favorite districts for later review.

4.5 Analytics and Visualization

An analytics dashboard was implemented to provide users with visual insights into the real-estate market. The dashboard displays information such as:

- Average prices by district
- Market growth rates
- Transaction activity
- Price comparison charts
- District performance indicators

Interactive charts and graphical representations were used to simplify data interpretation and improve decision-making. These visualizations allow users to compare different districts and analyze market trends more effectively.

4.6 GIS and Map Visualization

GIS capabilities use web mapping APIs (Google Maps on the predict page; Mapbox where configured). Users draw a rectangular area on the map to define the search region. District locations appear as markers; optional district labels improve orientation.

Maps connect analytical results (price, score, growth) to geographic context, so users see which neighborhoods match their budget and proximity preferences. Spatial selection, not manual district search, defines the recommendation scope.

4.7 Testing and Validation

Validation included **functional and integration testing**:

- **API health checks** for ML service and database connectivity.
- **End-to-end predict flows**: draw area, set preferences, run recommendation, verify ranked results and map markers.
- **Authentication**: registration, login, password reset, and favorites persistence.
- **Insights and transactions**: correct loading of charts and deal lists when backends are configured.
- **Error handling**: clear feedback when ML API, Supabase, or map keys are missing or invalid.

Model training used a **held-out test set** with metrics such as MAE, MAPE, and R^2 on predicted price per m^2 versus baseline medians. User testing indicated that the platform **simplifies district comparison** by combining map selection, ML-based pricing, and explainable ranked results in one workflow.

5. SYSTEM ARCHITECTURE AND IMPLEMENTATION

Robou is implemented as a three-tier web system for land valuation and district recommendation in the Eastern Province of Saudi Arabia (Dammam, Khobar, and Dhahran). The architecture separates user interaction, machine learning services, and persistent data/authentication so each layer can evolve independently

The system comprises:

5.1 Frontend layer

The **frontend layer** is the part the user sees and uses. It is built as a **Next.js** web application with **React**, responsible for pages, forms, bilingual text, and client-side map interaction.

Main responsibilities:

- **Presentation:** Home, predict (valuation), insights (market charts), transactions (deals list), account (profile and favorites), and authentication screens (login, signup, password recovery).
- **Localization:** Arabic and English via locale-based routes, including **right-to-left (RTL)** layout for Arabic.
- **User input:** On the predict page, users draw a **rectangular search area** on the map, choose residential or commercial land use, set a **budget range** (min/max price per m^2), and select **proximity to facilities** (near, medium, far).
- **Display of results:** Ranked district cards, estimated price per m^2, growth trend, services level, confidence, and short reasons; markers on the map for the top districts.
- **Session awareness:** Login state is managed with **Supabase Auth** in the browser; some actions (e.g., running a full recommendation or saving favorites) require an authenticated user.

5.2 Backend and ML layer

The **backend and ML layer** provides the **business logic and intelligence** of Robou. It is implemented as a **FastAPI** service (Python), running independently from the web UI

Main responsibilities:

- **REST APIs** for:
  - **District recommendation** inside a geographic area (with multi-criteria scoring: price, growth, services).
  - **Best areas** inside a user-drawn rectangle.
  - **Market insights** and **transaction/deal** listings for analytics pages.
  - **Health check** to verify the service and database connection.
- **Model loading at startup:** A pre-trained **aggregated price model** (gradient boosting on district–quarter features) and supporting feature tables (quarterly medians, growth rates, district coordinates, services indices).
- **Prediction pipeline:** For each candidate district—filter by location and city, predict **price per m^2** for the latest quarter, apply budget filters, compute a **composite score (0–100)**, rank districts, and attach **explainable labels** (reasons, growth trend, services level).
- **Data access for ML:** Reads **aggregated market data** from **MySQL** (districts, quarterly aggregates, growth, historical sales) when endpoints need live DB access.

5.3 Database and authentication layer :

The data layer is split intentionally between two systems, because Robou stores two different kinds of information: anonymous market analytics vs personal user data.

MySQL (relational database)

- Stores district master data, quarterly price aggregates, year-over-year growth, real sale records, and model version metadata.
- Used primarily by the ML backend for training pipelines and for endpoints that query live aggregates.
- Supports heavy analytical queries and offline data loading from preprocessed market datasets.

Supabase (PostgreSQL + Auth)

- Authentication: Email/password registration, login, password reset (including secure callback flow).
- User-specific data: Profiles and favorites (saved districts with predicted price per m^2), protected by row-level security so each user sees only their own records.
- Optional analytics mirror: Aggregate tables can be uploaded to Supabase so the frontend can serve insights and deals from the cloud when the ML API is not used alone.

5.4 Map visualization :

Map visualization is a cross-cutting capability used mainly in the presentation tier but relying on external geospatial services.

Main responsibilities:

- Interactive map on the predict page: pan, zoom, and draw a rectangular search region to define where recommendations apply.
- Spatial feedback: Show district markers for recommended areas; optional district name labels for orientation.
- Coordinate handling: Convert map clicks into geographic bounds sent to the backend; display returned centroids on the map.

Technologies:
Google Maps JavaScript API (via a React map library) on the predict flow; Mapbox (or related map libraries) may be used elsewhere for geocoding or alternate map views depending on configuration.

Relation to other layers:
Maps run in the browser; they do not store market data. Selected bounds are passed to the application layer, which forwards them to the ML backend for district filtering. Map API keys are configured in environment variables and restricted by domain in the map provider's console.

Table 1. Evaluation metrics of the final CatBoostRegressor model on the held-out test dataset using optimum density thresholds.

Model	Best Settings	R^2	MAE	MAPE
CatBoostRegressor	Min_deals = 10	0.645	359.9 SR/m^2	20.77%

6. RESULTS AND DISCUSSION

The evaluation of the proposed Robou platform demonstrates that the system successfully moves beyond providing an opaque price prediction to deliver actionable, explainable, and multi-faceted outputs designed for both expert and non-expert users. The practical utility of the platform is reflected in five core components generated during the inference phase:

- **District-Level Price Estimation (SR/m²):** predicted from the aggregated model using city, district, property type, and latest quarter statistics.
- **Ranked Top Districts (1–3):** within the user-drawn area, respecting budget and proximity preferences.
- **Explainable Metric Labels:** complete with real-time analytics regarding localized growth trends, services level, confidence tiers, and two natural-language reasons automatically generated in Arabic to justify the selection.
- **Geospatial Visualization:** projecting interactive map markers and district boundaries to facilitate efficient spatial comparison directly within the browser interface.
- **User Account Persistence:** enabling authenticated users to seamlessly bookmark, star, and track specific districts for longitudinal market review.

To maximize the statistical reliability of these outputs, extensive empirical experiments were conducted to evaluate the machine learning engine under various data density scenarios. Setting an appropriate data filtering threshold (min_deals) per district-quarter proved to be critical for model stabilization. Lower thresholds (e.g., min_deals < 5) introduce sparse, highly volatile transaction records that generate significant noise, causing the model to overfit. Conversely, setting the threshold too high (e.g., min_deals > 20) excessively limits the geographical scope of the platform, discarding developing neighborhoods and diminishing the localized granularity of the recommendations.

The optimal balance between geographical coverage and prediction accuracy was achieved using the CatBoostRegressor algorithm with the transaction threshold fixed at a minimum of 10 deals per district-quarter. Under this scenario, the model attained a test Coefficient of Determination (R^2) of 0.645, a Mean Absolute Error (MAE) of 359.90 SR/m², and a Mean Absolute Percentage Error (MAPE) of 20.77%, as detailed below in Table 1.

An MAE of 359.90 SR/m² indicates that the model's price predictions deviate by an average of roughly 360 SR from actual transaction medians. Given the steep price gradients and high spatial variance between premium coastal zones and expanding inland sectors across Dammam, Khobar, and Dhahran, this margin of error is highly acceptable. It confirms that the gradient boosting architecture effectively captured the underlying socio-economic and spatial features driving real estate dynamics in the Eastern Province, providing a robust empirical foundation for the subsequent multi-criteria ranking engine.

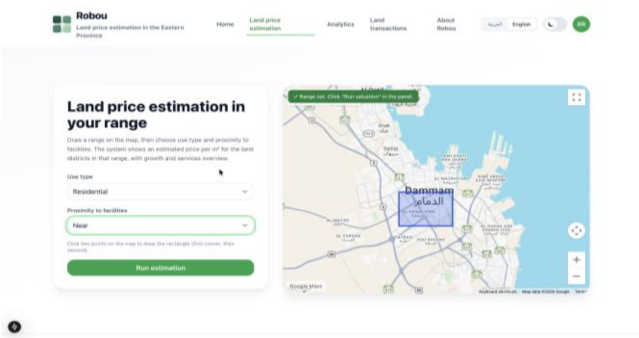

Figure 1. Robou interactive user interface showcasing map-based bounding box selection over Al-Faisaliyah district in Dammam, customized by land use type and facility proximity preferences.

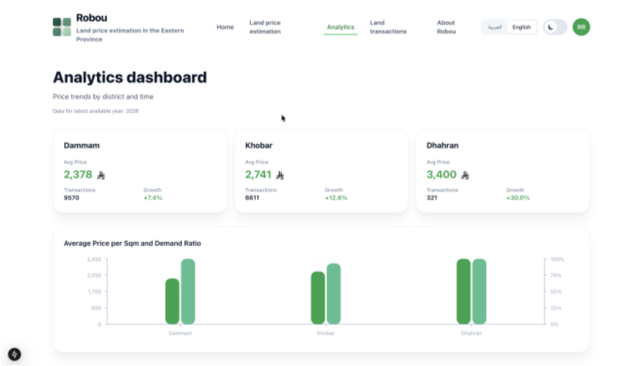

Figure 2. The market insights analytics dashboard displaying macro-level indicators across Dammam, Khobar, and Dhahran, including calculated average prices per square meter, overall transaction counts, and regional growth metrics.

7. CONCLUSION

This project presented Robou, a web-based decision-support system for land price estimation and district recommendation in the Eastern Province. By combining an interactive map-based frontend, a FastAPI ML backend with an aggregated price model, and Supabase-backed authentication and favorites, the system helps users explore where to buy land under budget and proximity constraints—not only how much a parcel might cost.

Main contributions:

- Integrated geospatial UI (bounding box selection) with multi-criteria district ranking (price, growth, services).
- Explainable outputs (scores, reasons, trends) suitable for non-expert users.
- Bilingual interface and alignment with local market context (Arabic districts, SAR pricing).
- Modular architecture separating user data (Supabase) from analytics data (MySQL/CSVs).

8. ACKNOWLEDGMENTS

The authors would like to express their appreciation to the project supervisor and the college of computer science and information technology for their support and guidance throughout the development of this project.

9. REFERENCES

[1] Abdrabo, A., Schlimmer, S., Ali, M., El Sherif, N., and Tempra, O. 2024. Saudi Arabia Land Sector Assessment. Background Paper. Arab Land Initiative. URL: https://arablandinitiative.gltn.net/sites/default/files/2024-10/docs/gulf-edit_saudi-arabia-land-sector_background-paper_10.7.2024_.pdf

[2] Algahtani, M. 2022. Constructing a House Price Index for Saudi Arabia. Journal of Real Estate Research. URL: https://www.tandfonline.com/doi/full/10.1080/10835547.2022.2105530

[3] Cultural and Environmental Studies Journal. 2024. Geographical Evaluation of Real Estate Services Offices in Abha. Research Article. URL: https://esiculture.com/index.php/esiculture/article/view/1114

[4] Alzamil, W. and Mohammed, A. 2018. Using Mobile GIS Technique in Real Estate Sector in Saudi Arabia: Reality & Challenges. ResearchGate Technical Report. URL: https://www.researchgate.net/publication/325113423

[5] Alzain, E., Alshebami, A., Aldhyani, T., and Alsubari, S. 2022. Application of AI for Predicting Real Estate Prices: Saudi Arabia. Electronics Journal (MDPI). URL: https://www.mdpi.com/2079-9292/11/21/3448

[6] Mathotaarachchi, R., Hasan, R., and Mahmood, A. 2024. Advanced ML Techniques for Predictive Modeling of Property Prices. Information Journal (MDPI). URL: https://www.mdpi.com/2078-2489/15/6/295

[7] Saudi Housing Program. 2024. Annual Report 2024. Saudi Vision 2030 Official Publication. URL: https://www.vision2030.gov.sa/media/vvyplaue/housing_program_annual_report_2024_en.pdf

[8] Knight Frank. 2023. The Saudi Report 2023. Industry Market Report. URL: https://content.knightfrank.com/research/2631/documents/en/the-saudi-report-2023-10029.pdf

[9] PwC. 2024. Real Estate Outlook 2024: Emerging Trends in Real Estate. PricewaterhouseCoopers Global Report. URL: https://www.pwc.com/gx/en/asset-management/emerging-trends-real-estate/assets/uli-emerging-trends-global-report-2024.pdf