BorBann
A Real Estate Information Platform
(Project Proposal)

by

Pattadon Loyprasert 6510545608
Sirin Phungkun 6510545730

Submitted as
Software Requirements Specification
for
01219395 Innovative Software Group Project Preparation

Faculty of Engineering
Kasetsart University
Bangkok, Thailand

March 2025

BorBann
A Real Estate Information Platform
(Project Proposal)

by

Pattadon Loyprasert 6510545608
Sirin Phungkun 6510545730

This Project Submitted in Partial Fulfillment of the Requirements
for Bachelor Degree of Engineering
(Software Engineering)

Department of Computer Engineering, Faculty of Engineering
KASETSART UNIVERSITY
Academic Year 2025

Approved by:

Advisor ..... Date.../.../...
(Assoc.Prof.Dr. Kitsana Waiyamai)

Table of Contents

List of Tables **v**

1 Introduction **1**

1.1 Background 1

1.2 Problem Statement 1

1.3 Solution Overview 2

1.4 Target User 4

1.5 Benefit 4

1.6 Terminology 5

2 Literature Review and Related Work **6**

2.1 Competitor Analysis 6

3 Requirement Analysis **8**

3.1 Stakeholder Analysis 8

3.2 User Stories 10

3.3 Use Case Diagram 17

3.4 Use Case Model 19

3.5 User Interface Design 22

4 Software Architecture Design **36**

4.1 Sequence Diagram 36

4.2 AI Component 37

References **43**

List of Tables

1.1 Target Users and Their Needs 4
1.2 Terminology 5

2.1 Feature Comparison: BorBann vs Other Platforms 6
2.2 Comprehensive Technical Comparison 7

3.1 BorBann Platform Stakeholder Summary 8
3.2 User Stories with Acceptance Criteria 11
3.3 View Property Insights Use Case 20
3.4 Explainable Price Prediction Use Case 21

Chapter 1

Introduction

1.1 Background

The global real estate market shows positive growth trends despite various challenges. Real estate market remains a key sector for people like homebuyers and investors, influenced by complex characteristics with heterogeneous nature and affected by numerous elements like policies and consumer trends. The market is projected to grow from $4,143.71 billion in 2024 to $4,466.58 billion in 2025, and global real estate investment volumes continue to increase in 2024[1].

Thailand's real estate market faces both challenges and opportunities in the current economic climate. The real estate market in Thailand, especially in Bangkok in 2024-2025, faces challenges such as high household debt, strict credit conditions, and rising costs causing market contraction, decreasing project launches by 43.72% in 2024[2]. However, it also presents opportunities for recovery through foreign investment driven by tourism recovery, government incentives, and technological advancements. Condominiums remain attractive for rental yields and foreign buyers, while the housing market has moderate growth supported by infrastructure development.

Crucially, these market opportunities can significantly undermined by Thailand's real estate information ecosystem, which suffers from several structural limitations, including outdated listings, lack of official transaction data, and data fragmentation[3]. Addressing these issues is essential for maximizing Thailand's market potential and enabling effective decision-making for both homebuyers and investors seeking to capitalize on the available opportunities.

1.2 Problem Statement

Even with large amounts of real estate data available through various channels including property listings, historical transaction records, and news, investors still face challenges in aggregating, contextualizing, and deriving action from these sources. Additionally, the Thai real estate market has different characteristics from other countries because of the

limited availability of official property transaction records, listing duplication across platforms, and less standardized data compared to more mature markets.

Current platforms in Thailand like DDProperty, Hipflat, and Baania lack many important contextual factors such as climate risk assessments, neighborhood-specific news, and local beliefs that influence property valuation and real estate investment in the long term.

While platforms like Zillow and House Canary have advanced and comprehensive real estate analytics, predictive analytics, and user-friendly interfaces, they do not operate in Thailand. Even if these advanced platforms were to enter the Thai market, they would face challenges adapting their algorithms to the local context. Their models are calibrated to markets with standardized property classifications, valuations that vary by developers, and consistent transaction data—elements that are limited in Thailand's real estate ecosystem.

The BorBann platform addresses these challenges and aims to help users by creating a real estate data platform integrated with artificial intelligence, geospatial analytics, and data aggregation by focusing on analytics rather than transaction facilitation.

1.3 Solution Overview

BorBann will function as real estate data platform to users, integrating multiple data sources with advanced analytics. Below are the features and their details.

Customizable Automated Data Integration Pipeline

- **Automated schema inference:** Analyze website structures to identify and extract key data elements
- **Field mapping:** Recognize equivalent fields across different sources (e.g., "price" vs "cost")
- **Integration framework:** Seamless connection with data export systems
- **Multi-source support:** Process data from websites, APIs, and uploaded files

Retrain Model with Data from Pipeline

- **Custom prediction model:** Create custom prediction models by combining their pipeline data with platform data sources

Local Contextual Analytics

- **Environmental risk assessment:** Evaluate flood risk, natural disaster vulnerability, and air quality
- **Facility proximity analysis:** Calculate accessibility to schools, hospitals, transit, and commercial centers

- **Neighborhood quality scoring:** Generate composite metrics for area evaluation

Explainable Price Prediction Model

- **Feature importance analysis:** Quantify and rank factors influencing property prices
- **Adjustable characteristics modeling:** Adjust property characteristics to visualize price impacts
- **Confidence intervals:** Provide lower/upper price bounds for realistic expectations
- **Factor categorization:** Group influences by type (property features, location, market trends)
- **Natural language explanations:** Generate readable summaries of price determinants
- **Visual breakdowns:** Display contribution percentages and relationship graphs

Geospatial Visualization

- **Heatmap generation:** Create density visualizations for environmental factors, pricing, and metrics
- **Geospatial analytics:** Calculate analytics for custom geographic areas

1.4 Target User

User Type	Description	Needs
Real Estate Investors	Individuals focused on maximizing long-term investment, including foreign investors	• Investment analysis • Supporting data for decision making
Homebuyers	First-time purchasers, residents looking to relocate within Thailand, and expats seeking housing	• Property comparisons • Neighborhood insights • Pricing guidance

Table 1.1: Target Users and Their Needs

1.5 Benefit

The BorBann platform will provide numerous benefits to the Thai real estate market. It improves market transparency by enhancing accessibility to market information. It also helps both homebuyers and investors reduce research time, achieve lower transaction risks, and discover overlooked investment opportunities. Additionally, the platform will effectively represent the unique characteristics of the Thai real estate market.

1.6 Terminology

Term	Definition
Local Analytics	Analysis focused on extremely specific geographic areas, such as neighborhoods or even individual streets, to provide highly relevant insights.
Price Prediction Model	An algorithm or statistical model that forecasts property values based on historical data, market trends, and various property characteristics.
Proximity Analysis	The study of spatial relationships between geographic features, typically to evaluate the distance between properties and amenities or services.
Geospatial Visualization	The graphical representation of data with a geographic or spatial component, often through maps and interactive displays.

Table 1.2: Terminology

Chapter 2

Literature Review and Related Work

This chapter presents a review of existing platforms in the domain of real estate analytics and information systems. This review includes both international and Thai platforms, their features, strengths, and limitations. The analysis establishes the current state of real estate information platforms and identifies opportunities for the BorBann platform to address unmet needs in the Thai market.

2.1 Competitor Analysis

Many real estate platforms primarily function as property listing aggregators rather than comprehensive information systems. Their features support transaction facilitation through showcasing available properties, while offering limited analytical tools for market understanding. However, platforms like House Canary represent exceptions, operating specifically as information systems that provide investors with data analytics and market insights to support evidence-based decision-making in real estate investments.

Feature	BorBann (Proposed)	DDProperty	Hipflat	House Canary	Zillow
Customizable Automated Data Integration Pipeline	**Yes**	No	No	No	No
Retrain Model with Data from Pipeline	**Yes**	No	No	No	No
Local Contextual Analytics	**Yes**	No	No	Yes (Not optimized for Thailand)	Yes (Not optimized for Thailand)
Explainable Price Prediction Model	**Yes**	No	No	No	No
Geospatial Visualization	**Yes**	Yes	Yes	Yes	Yes

Table 2.1: Feature Comparison: BorBann vs Other Platforms

Table 2.1 demonstrates BorBann's technical advantages in the real estate analytics market. While all platforms offer geospatial visualization, BorBann's implementation includes advanced analytics like climate assessment, matching international platforms but

surpassing local Thai competitors. BorBann's automated data integration pipeline collects analytics-ready data automatically, unlike Thai platforms that rely on user inputs. Also, user can use that data to create their custom models. For local contextual analytics, BorBann provides Thailand-optimized insights including weather patterns and population density, whereas DDProperty/Hipflat only show basic nearby facilities, and international platforms lack Thailand-specific optimization. Most distinctively, BorBann's price prediction model prioritizes explainability and interpretability, revealing the reasoning behind valuations rather than presenting opaque predictions like competing platforms.

Technical Aspect	DDProperty	Hipflat	House Canary	Zillow	BorBann (Proposed)
Data Sources	User-submitted, Proprietary	User-submitted, Proprietary	Proprietary	Multiple Sources, Proprietary	Open Data, User-submitted
ML Implementation	Basic Prediction	None	Black-box Models	Black-box Models	Explainable Models

Table 2.2: Comprehensive Technical Comparison

Table 2.2 highlights key technical differences between BorBann and competing platforms. For data sources, while competitors rely heavily on government or user-submitted data, BorBann uniquely leverages open data combined with APIs to build a more comprehensive dataset. Regarding machine learning, BorBann distinguishes itself by implementing explainable models, providing transparency in its predictions. This contrasts with DDProperty's basic prediction capabilities, Hipflat's complete lack of ML features, and the black-box approaches of House Canary and Zillow where prediction logic remains hidden from users.

Chapter 3

Requirement Analysis

3.1 Stakeholder Analysis

The stakeholder landscape for BorBann includes primary stakeholders who directly interact with the platform and secondary stakeholders who indirectly influence its adoption and effectiveness. Recognizing these groups' specific needs helps develop a platform that delivers value across the real estate ecosystem.

Stakeholder Category	Stakeholder Group	Key Interests	Primary Requirements
Primary	Real Estate Investors	Environment risk assessment	Advanced analytics, predictive modeling, risk scoring
Primary	Homebuyers	Affordability, area quality, lifestyle alignment, environmental factors	User-friendly interface, neighborhood insights, price comparisons, risk assessments
Secondary	Real Estate Agencies	Market positioning, client advisement, property valuation	Property data, reliable market insights

Table 3.1: BorBann Platform Stakeholder Summary

The primary stakeholders are direct users who rely on BorBann's capabilities for informed real estate decisions. Their requirements range from investment analytics to intuitive neighborhood quality assessments. Secondary stakeholders like real estate agencies can influence platform adoption and serve as valuable data partners.

Primary Stakeholders

Primary stakeholders are those who directly use or are significantly affected by the BorBann platform. They rely on its capabilities to make informed decisions about real estate investments, purchases, and market trends.

Real Estate Investors

Real estate investors use BorBann to identify profitable investment, assess risks, and make data-driven decisions.

Interests and Concerns:

- Maximizing return on investment
- Identifying growth areas
- Risk assessment
- Diversification across neighborhoods and property classes

Requirements:

- Advanced analytics tools that is customizable
- Predictive price modeling with easy to understand explaination
- Risk scores in each area based on multiple factors

Homebuyers

Homebuyers rely on BorBann to find properties that align with their budget, lifestyle, and long-term needs. The platform helps them assess affordability, neighborhood quality, and potential risks.

Interests and Concerns:

- Property affordability and financing options
- Area quality
- Location amenities and lifestyle alignment
- Environmental factors including flood and pollution risks
- Commute time to work, schools, and essential services
- School districts and educational quality

Requirements:

- Intuitive, user-friendly interface with minimal learning curve
- Comprehensive neighborhood insights including safety metrics
- Price comparison tools with historical context

- Property quality and developer reputation metrics
- Detailed flood risk and environmental quality assessment
- Transit accessibility maps with time-based visualizations
- Education quality indicators and school zone mapping

Secondary Stakeholders

Secondary stakeholders are indirectly impacted by the BorBann platform, influencing its adoption, data availability, and overall market reach.

Real Estate Agencies

Real estate agencies act as intermediaries between property buyers and sellers. They use BorBann to improve their advisory services, gain a competitive edge, and provide market insights to clients.

Interests and Concerns:

- Market positioning relative to competitors
- Client advisement based on reliable data
- Access to comprehensive property information
- Accurate property valuation to support transactions

Influence:

- Potential data partners and platform promoters
- Can significantly influence adoption rates among clients
- May provide valuable transaction data not available elsewhere

3.2 User Stories

User stories capture the essential needs and goals of the BorBann platform's target users from their perspective. These stories follow the standard format: "As a [user type], I want to [action/feature] so that [benefit/value]."

Table 3.2: User Stories with Acceptance Criteria

User Story	Acceptance Criteria
Customizable Automated Data Integration Pipeline	
As a non-technical user, I want to input a website URL and have the system automatically generate a scraping configuration, so I can collect data without coding skills	• System provides visual confirmation of detected data structure • Non-technical users can successfully create a working pipeline in under 5 minutes
As a user, I want to paste multiple URLs from the same website pattern and have the system recognize the common structure, so I can efficiently collect data from similar pages	• System identifies common patterns across multiple URLs from the same website • A single configuration works for all provided URLs from the same pattern
As a user, I want to upload data files (CSV, JSON) as alternative data sources to integrate with my scraped data	• System accepts uploads of CSV and JSON files up to 50MB • Automatic schema detection for uploaded files with 90% accuracy
As a user, I want to customize the output format and template via an intuitive UI	• User can select from at least 4 output formats (JSON, CSV, SQLite, YAML) • UI provides visual preview of output format before confirmation
As a user, I want to schedule my data pipeline to run automatically at specified intervals	• Interface allows setting schedule frequency (hourly, daily, weekly, monthly) • Timezone selection is available for scheduling

Continued on next page

Table 3.2 – Continued from previous page

User Story	Acceptance Criteria
As a user, I want a dashboard showing all my data pipelines with their status and last run time	• Dashboard displays all pipelines with status indicators • Last run time and next scheduled run are clearly shown
Local Contextual Analytics	
As a user, I can see all contextual data for specific areas to make informed decisions	• System displays at least 5 contextual data points for any selected area • Historical trends for environmental factors available for at least 12 months
As a user, I want to see flood risk assessments for properties with historical flooding data	• Flood risk presented on a 5-point scale with historical context • System shows flood history for the past 10 years when available
As a user, I want to see daily air quality metrics around a property with historical trends	• Air quality index displayed with daily updates • Historical air quality data presented in trend graphs
As a user, I want to see all schools within a custom radius, including distance, ratings, and types	• System displays all schools within user-defined radius • Each school listing includes distance, type, and quality rating

Continued on next page

Table 3.2 – Continued from previous page

User Story	Acceptance Criteria
As a user, I want to view healthcare facilities near a property with distance and service information	• Healthcare facilities categorized by type (hospital, clinic, etc.) • Distance and basic service information provided for each facility
Explainable Price Prediction Model	
As a user, I want to see how specific contextual factors influence the property's predicted price	• System displays top 5 factors influencing price with percentage contribution • Visual indicators show positive/negative impact of each factor
As a user, I want the model to be interpretable so I can understand factors affecting the price	• Plain language explanations accompany each prediction • Interactive elements allow exploration of factor relationships
As a user, I want a statement or reason to back the prediction so I can trust the system's valuation	• Each prediction includes at least 3 specific supporting statements • System indicates confidence level for each prediction
As a user, I want to see a predicted price range for any property I select	• System shows lower and upper bounds for predicted price
As a user, I want to understand how the model derives the result so I can explain the valuation to others	• System provides visual breakdown of prediction process
Geospatial Visualization	

Continued on next page

Table 3.2 – Continued from previous page

User Story	Acceptance Criteria
As a user, I can see property listings on the map and click them to view detailed information	• Property details popup appears within 1 second of clicking a marker • Popup contains at least 5 key property attributes
As a user, I can see sections of the same property group to identify properties from the same development	• Properties from same development visually grouped with distinct boundaries • Group name appears when hovering over grouped properties
As a user, I can see multiple map visualization types to analyze different environmental factors	• System offers at least 5 different visualization overlays • Users can toggle between visualizations without page reload
As a user, I want to pan and zoom on an interactive property map to explore different areas efficiently	• Map responds to standard pan/zoom gestures within 100ms • Property markers adjust density based on zoom level
As a user, I want to set a custom radius around a point on the map to analyze the surrounding area	• User can place and adjust analysis radius on any map location • Contextual analytics update in real-time as radius is moved
Retrain Model with Data from Pipeline	

Continued on next page

Table 3.2 – Continued from previous page

User Story	Acceptance Criteria
As a non-technical user, I want to select one of my existing data pipelines as a source for model training	• User can select any active pipeline as a data source through a simple dropdown • System validates data compatibility before starting training process
As a user, I want to select from recommended model types appropriate for my data	• System suggests optimal model types based on data characteristics • Each model type includes a simple explanation of its strengths and use cases
As a user, I want to start the model training process with a single click after selecting my data sources	• Training begins with a single action after configuration • System provides confirmation that training has started
As a user, I want to see how accurate my trained model is compared to platform default models	• Performance metrics displayed with comparative benchmark against standard models • Visualizations show improvement or differences in prediction accuracy
As a user, I want to activate my newly trained model with a single click to apply it across the platform	• Model activation changes system behavior immediately • Visual indicator shows which model is currently active

Continued on next page

Table 3.2 – Continued from previous page

User Story	Acceptance Criteria
As a user, I want to see a list of all models I've trained with performance metrics and creation dates	• Management interface displays all user models with key metadata • Models can be sorted and filtered by different attributes
As a user, I want to receive clear explanations if my pipeline data is unsuitable for training	• System provides specific feedback about data quality issues • Suggestions for data improvements are provided when problems are detected

3.3 Use Case Diagram

The Use Case Diagram for the BorBann platform, shown in Figure 3.1, illustrates the primary interactions between the system and its three main user types: Real Estate Investors, Homebuyers, and Property Developers. The diagram captures the core functionality of the platform and how different users interact with its features.

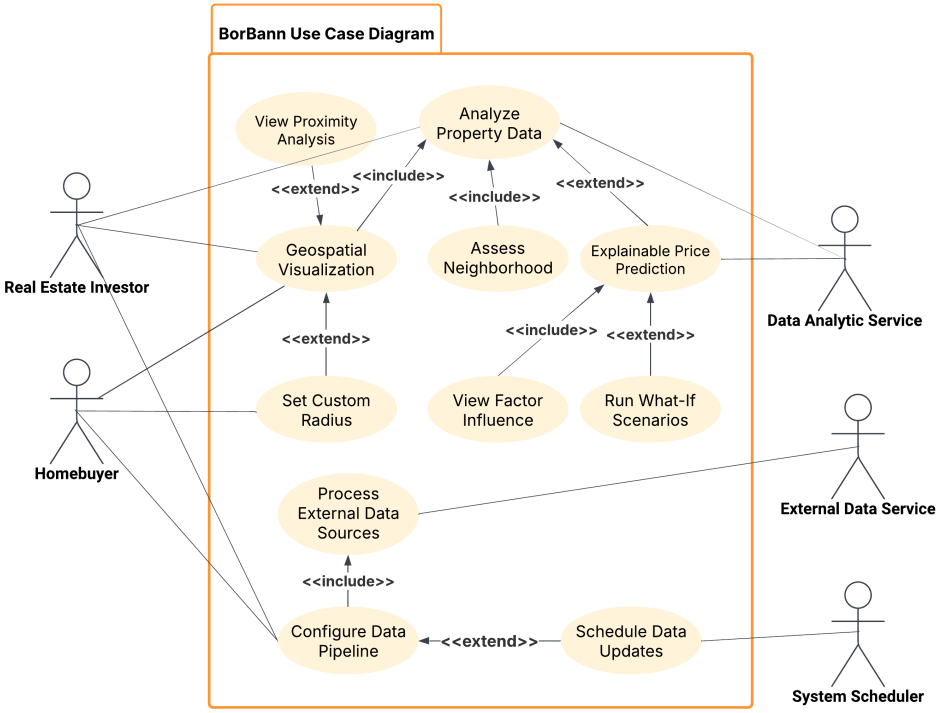

Figure 3.1: Use Case Diagram

Primary Actors

The use case diagram identifies five key actors interacting with the BorBann system:

- **Real Estate Investor** Seeks in-depth property analytics, market trends, and investment decision support.
- **Homebuyer** Interested in residential properties, lifestyle factors, and long-term value.
- **Data Analytic Service** Provides analytic capabilities to support prediction and data analysis features.
- **External Data Service** Supplies third-party data inputs via APIs and other integrations.

- **System Scheduler** Automates routine system tasks such as scheduled data refresh operations.

Core Use Cases

The BorBann platform provides the following primary functionalities:

- **Analyze Property Data** Allows users to examine detailed property information, metrics, and comparisons.
  - *Includes:* Assess Neighborhood
  - *Extended by:* Explainable Price Prediction
- **Geospatial Visualization** Interactive map-based visualization of property and neighborhood data.
  - *Extended by:* Set Custom Radius, View Proximity Analysis
- **Assess Neighborhood** Evaluates contextual factors around properties (e.g., schools, amenities).
  - *Includes:* View Factor Influence
  - *Extended by:* Check Environmental Risks
- **Explainable Price Prediction** AI-generated pricing insights with interpretable influencing factors.
  - *Includes:* View Factor Influence
- **Configure Data Pipeline** Enables setup of automated ingestion of external data from APIs and sources.
  - *Includes:* Process External Data Sources
  - *Extended by:* Schedule Data Updates

Extended Functionality

The following optional use cases extend the platform's functionality:

- **View Proximity Analysis** Triggered when users want detailed proximity-based data (e.g., nearby schools, transport).
- **Set Custom Radius** Adds user-defined range settings for map-based filtering and analysis.
- **Schedule Data Updates** Automates periodic refreshes of data pipelines using system scheduler logic.
- **Check Environmental Risks** Enables users to access data on environmental threats (e.g., flood zones, pollution).

Actor-System Interactions

Each actor interacts with the system in distinct ways, as shown in the diagram:

- **Real Estate Investor**
  - Initiates *Analyze Property Data*, *Geospatial Visualization*, and *Configure Data Pipeline*
  - Benefits from extended insights like *Set Custom Radius*, *View Proximity Analysis*, and *Explainable Price Prediction*
- **Homebuyer**
  - Accesses *Analyze Property Data*, *Geospatial Visualization*, *Assess Neighborhood*, and *Explainable Price Prediction*
  - Makes use of neighborhood-specific features like *View Factor Influence* and *Check Environmental Risks*
- **Data Analytic Service**
  - Collaborates with the system to enable *Explainable Price Prediction*
- **External Data Service**
  - Supports *Process External Data Sources* as part of the data ingestion workflow
- **System Scheduler**
  - Triggers *Schedule Data Updates* as an automated background process

Use Case Relationships

- **<<include>>** Represents mandatory sub-functions:
  - *View Factor Influence* is included in both *Assess Neighborhood* and *Explainable Price Prediction*
  - *Process External Data Sources* is included in *Configure Data Pipeline*
- **<<extend>>** Represents optional or conditional behavior:
  - *Set Custom Radius* and *View Proximity Analysis* extend *Geospatial Visualization*
  - *Check Environmental Risks* extends *Assess Neighborhood*
  - *Schedule Data Updates* extends *Configure Data Pipeline*

3.4 Use Case Model

The Use Case Model details the interactions depicted in the Use Case Diagram, describing each use case within BorBann's core features.

Use Case Name	**View Property Insights**
Actors	Real Estate Investor, Homebuyer, Property Developer
Description	Provides users with comprehensive analytics and contextual information about specific properties
Preconditions	User has selected a property from search results
Basic Flow	1. User selects a property for detailed viewing 2. System retrieves comprehensive property data 3. System presents property details with analytics 4. System displays neighborhood characteristics 5. System shows historical performance metrics 6. User reviews information
Alternative Flows	• User can extend to save the property as a favorite • User can extend to view explainable price predictions • User can request additional specific analytics
Postconditions	User gains comprehensive insights about the property
Associated Feature	Local Contextual Analytics - providing trend analysis and neighborhood-specific insights

Table 3.3: View Property Insights Use Case

Use Case Name	**Explainable Price Prediction**
Actors	Real Estate Investor, Homebuyer
Description	Provides transparent, interpretable price predictions with detailed explanations of contributing factors
Preconditions	User is viewing property insights
Basic Flow	1. User requests price prediction for a property 2. System calculates predicted price using its models 3. System generates explanation of factors influencing the prediction 4. System presents prediction with confidence interval 5. System displays factor weights and their impact 6. User reviews prediction and explanation
Alternative Flows	• User can adjust factors to see impact on prediction • User can compare prediction with market averages • User can save prediction for future reference
Postconditions	User understands the predicted price and the reasoning behind it
Associated Feature	Explainable Price Prediction Model - providing transparent explanations for price predictions

Table 3.4: Explainable Price Prediction Use Case

3.5 User Interface Design

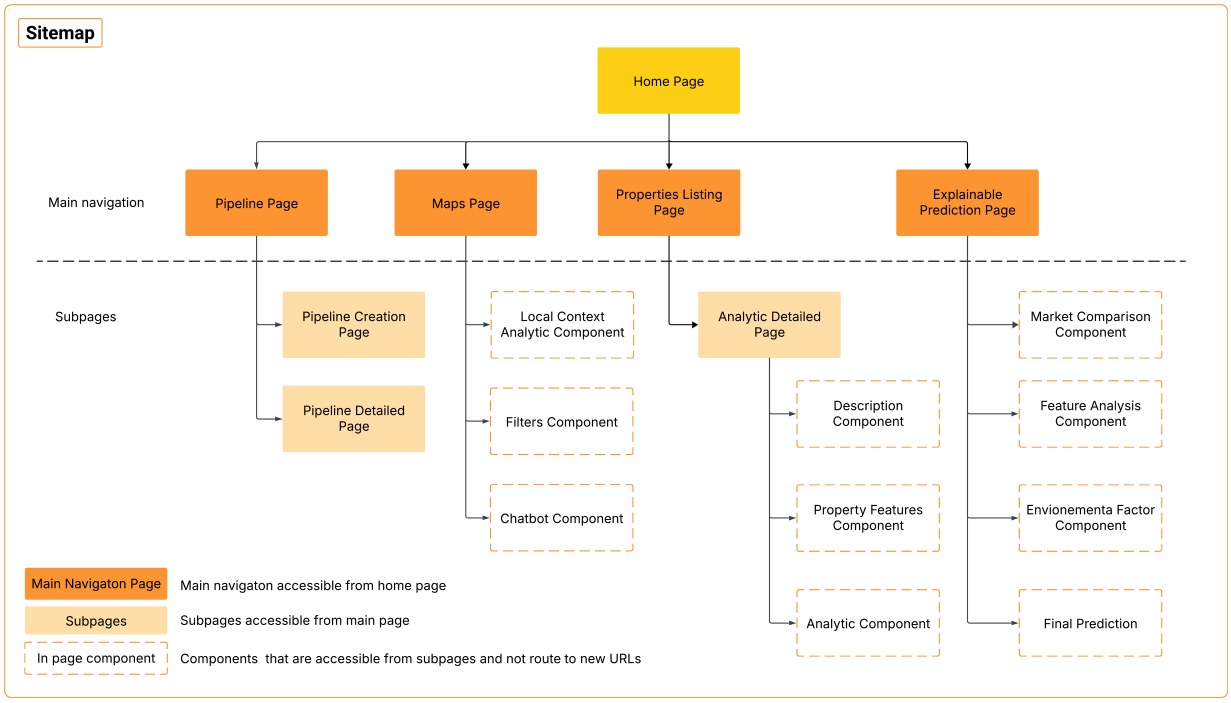

Figure 3.2: Sitemap - Platform Structure Overview

Figure 3.2 shows the structure of the BorBann platform. The Home Page acts as the central access point leading to four main sections: Pipeline, Maps, Properties Listing, and Explainable Prediction pages.

Each main section connects to specific subpages. The Pipeline section includes Creation and Detailed pages. The Maps section features Local Context Analytics, Filters, and Chatbot components. The Properties section provides Analytic Detailed pages with Description, Property Features, and Analytics components. The Explainable Prediction section offers Market Comparison, Feature Analysis, Environmental Factor analysis, and Final Prediction components.

This structure organizes the platform's key functions in a logical flow, making it easy for users to navigate between data management, visualization, and prediction features.

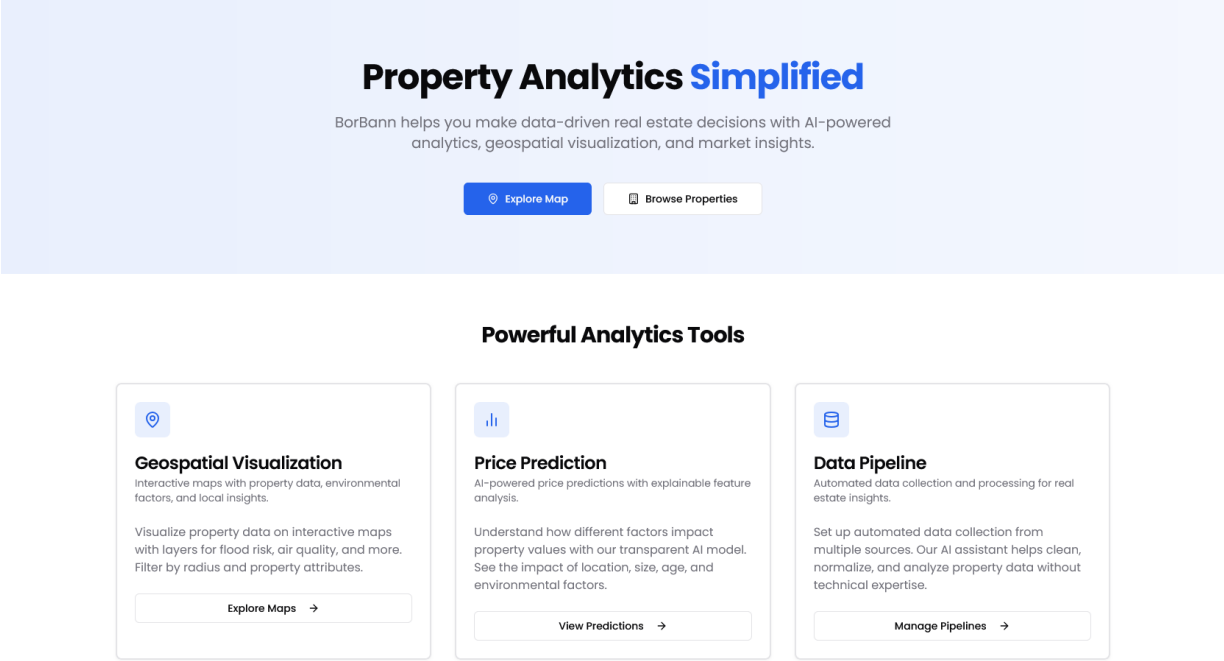

Figure 3.3: BorBann Platform Homepage

Figure 3.3 shows the BorBann platform homepage, which provides an intuitive entry point to the system. The page presents the core value proposition with "Property Analytics Simplified" and briefly explains how BorBann helps users make data-driven real estate decisions through analytics, geospatial visualization, and market insights. Two primary action buttons - "Explore Map" and "Browse Properties" - enable quick access to key functionality. The lower section showcases three main analytics tools: Geospatial Visualization for interactive property mapping, Price Prediction with explainable AI features, and Data Pipeline for automated data collection and processing.

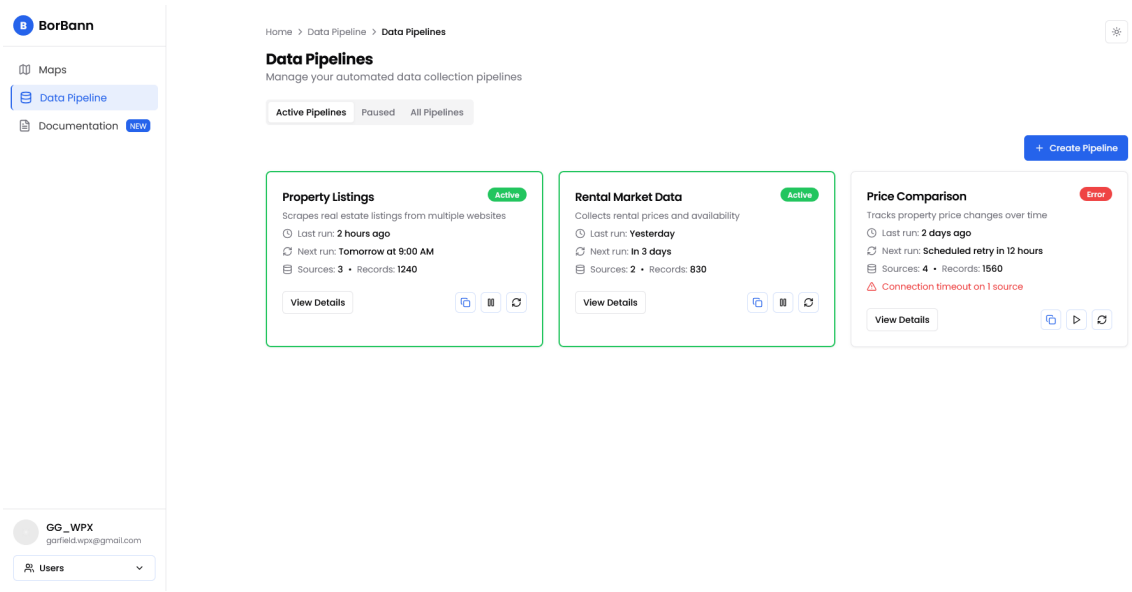

Figure 3.4: Data Pipeline Management Interface

Figure 3.4 showcases the Data Pipeline management dashboard, providing users with comprehensive control over their automated data collection processes. The interface features intuitive navigation with filtering options for Active, Paused, and All Pipelines, enabling efficient workflow management.

Each pipeline card presents essential operational metrics including last run timestamp, next scheduled execution, number of data sources, and total records processed. This at-a-glance view allows users to quickly assess data freshness and monitor collection performance across multiple pipelines.

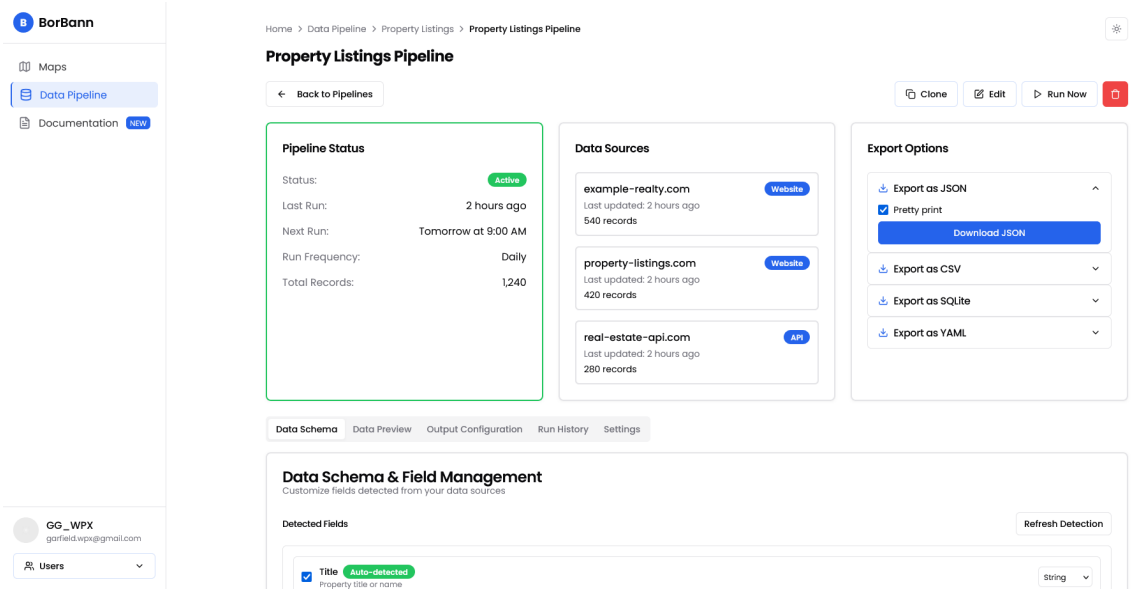

Figure 3.5: Data Integration Pipeline Detail - Overview

Figure 3.5 illustrates the pipeline detail view, highlighting key operational components including current status indicators, connected data sources, and available export formats. This overview provides users with immediate visibility into pipeline configuration and functionality.

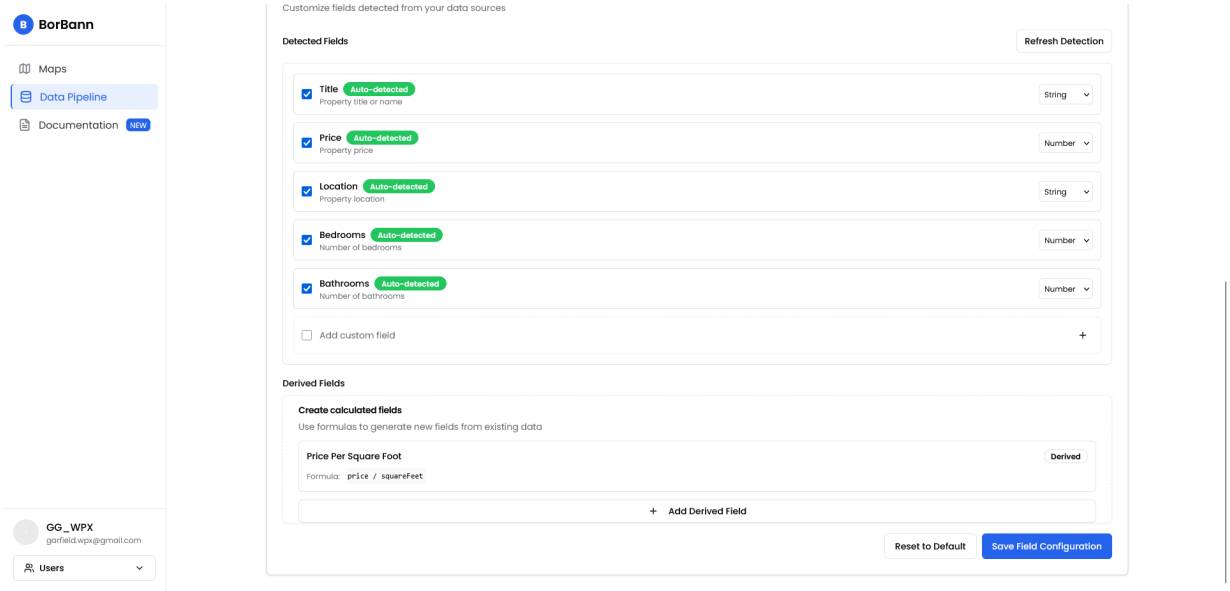

Figure 3.6: Data Integration Pipeline Detail - Field Management

Figure 3.6 displays the field management section of the pipeline detail page. This interface empowers users to customize their data structure by managing output fields and creating derived fields through a visual formula builder. Users can transform raw data into meaningful metrics without writing code.

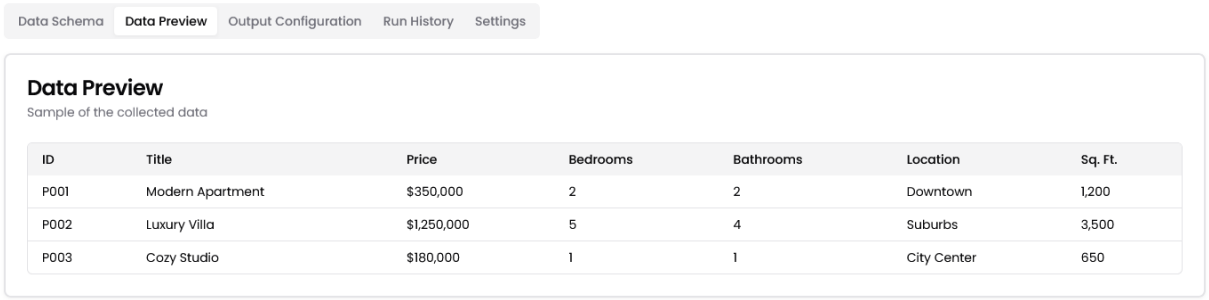

Figure 3.7: Data Integration Pipeline Detail - Output Data Preview

Figure 3.7 presents the data preview tab, where users can examine sample records generated by their pipeline. This real-time preview functionality allows users to validate data quality and structure before export or analysis.

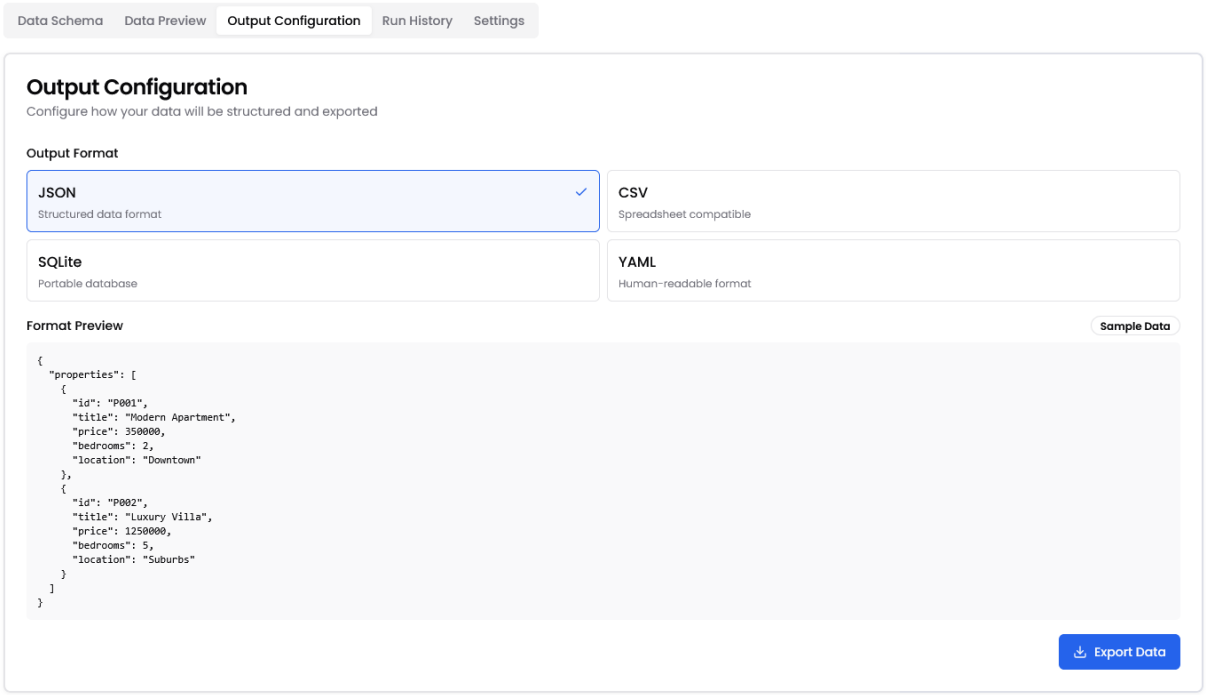

Figure 3.8: Data Integration Pipeline Detail - Export Configuration

Figure 3.8 shows the export configuration interface, where users can precisely define output schemas for their data exports. This tab enables users to select specific fields, customize formatting, and choose from multiple export formats to meet their downstream requirements.

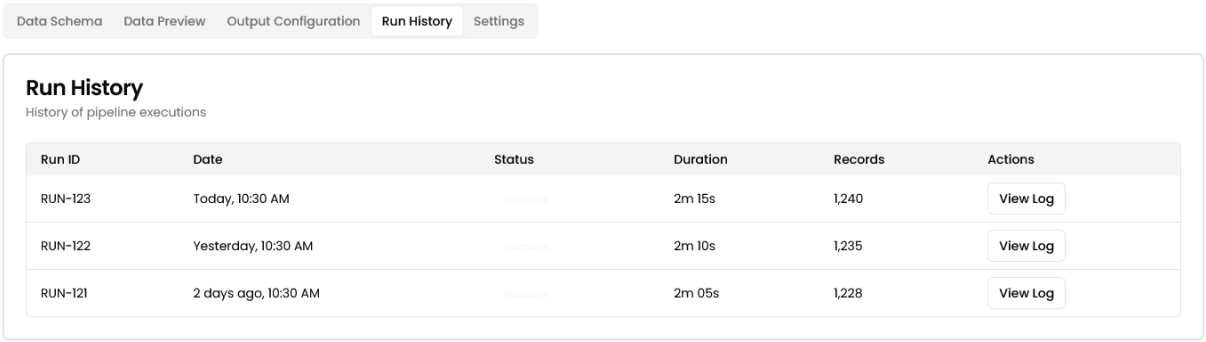

Figure 3.9: Data Integration Pipeline Detail - Run History

Figure 3.9 displays the run history tab, providing a chronological log of pipeline executions. This audit trail includes execution timestamps, duration metrics, and status indicators, enabling users to monitor performance trends and troubleshoot any execution issues.

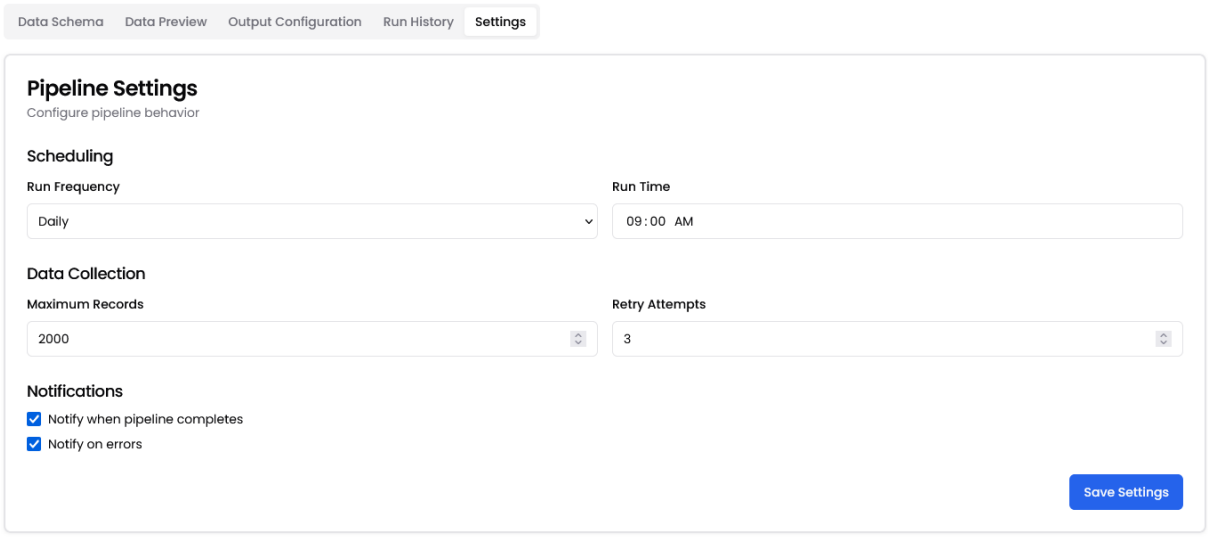

Figure 3.10: Data Integration Pipeline Detail - Pipeline Settings

Figure 3.10 illustrates the pipeline settings tab, where users can configure automation parameters including execution schedules, notification preferences, and retry policies. These controls allow users to establish reliable data collection routines that align with their operational requirements.

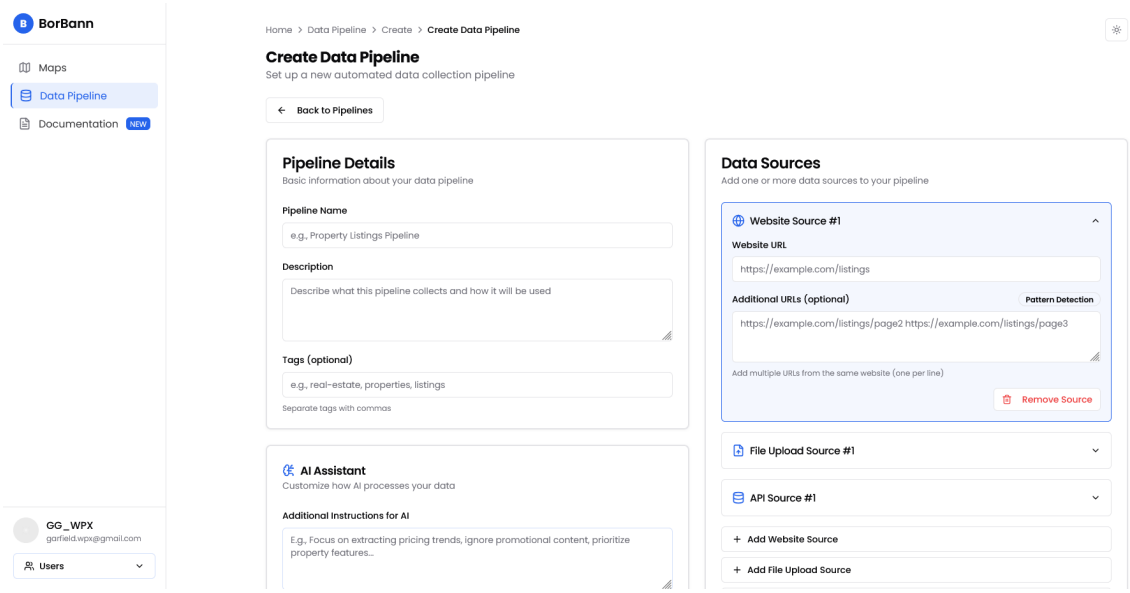

Figure 3.11: Data Integration Pipeline - Creation Interface

Figure 3.11 shows the pipeline creation interface where users input website URLs for automated data extraction. The form enables non-technical users to configure scraping operations without coding knowledge.

Figure 3.12 displays automation settings for pipeline execution with scheduling options and AI-assisted extraction configuration. Users can set run frequency and notification

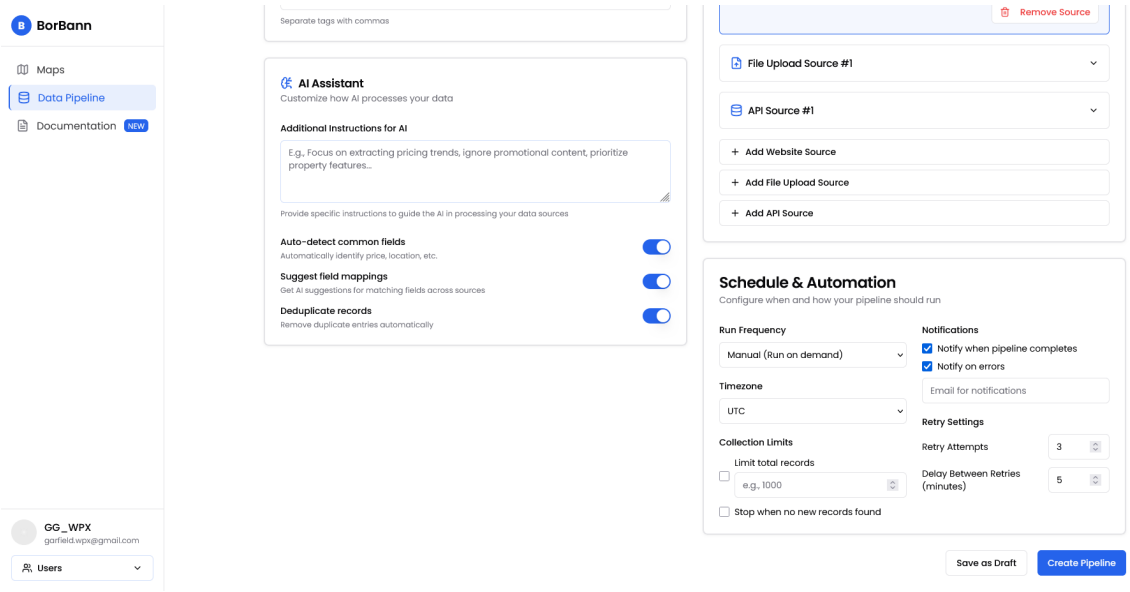

Figure 3.12: Data Integration Pipeline - Advanced Configuration and Scheduling

preferences to maintain automated data collection.

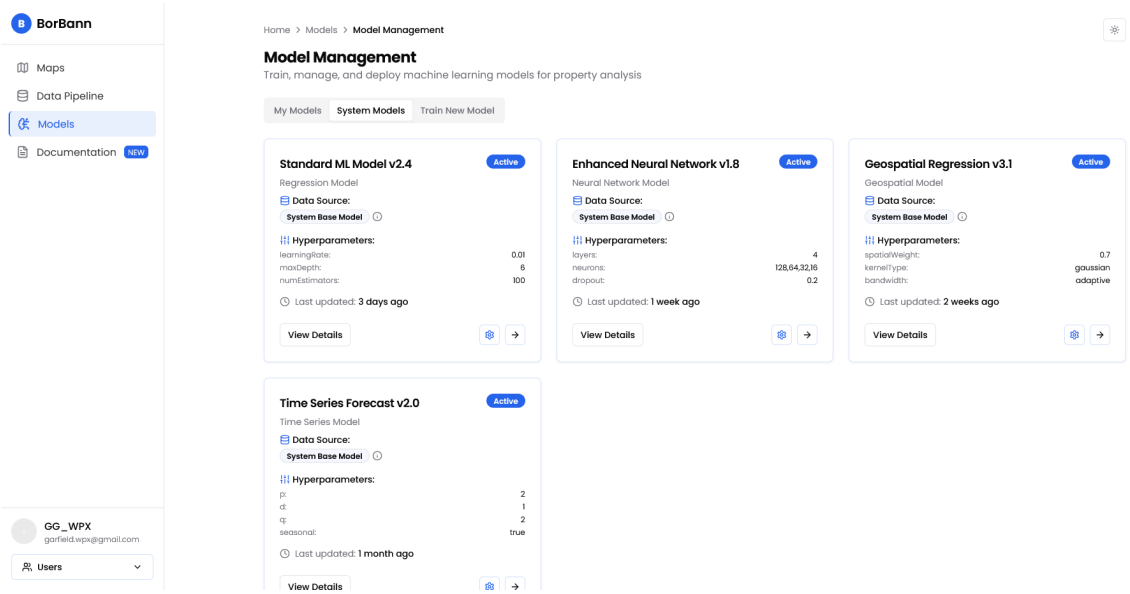

Figure 3.13: Model Management Dashboard

Figure 3.13 presents the model management dashboard where users can view and control prediction models. The interface displays performance metrics and allows single-click model activation.

Figure 3.14 shows the model creation interface for configuring new prediction models. Users can select data pipelines as training sources and choose from recommended algorithms with explanations of their strengths.

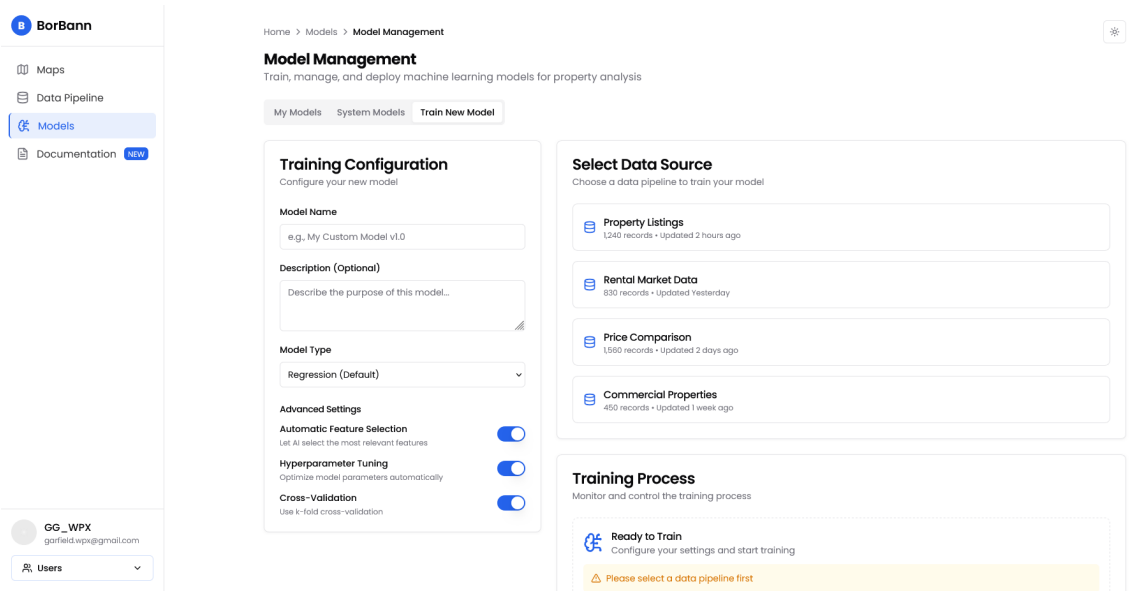

Figure 3.14: Model Creation Interface

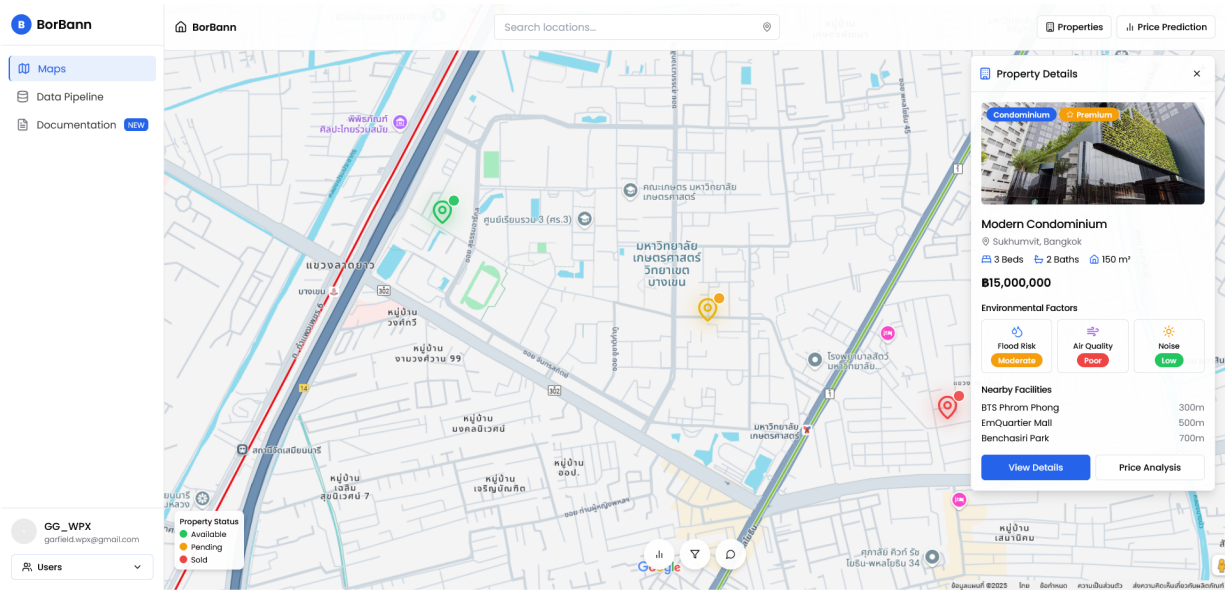

Figure 3.15: Geospatial Visualization - Main Map Interface

Figure 3.15 displays the interactive property map with color-coded markers and navigation controls. The interface supports pan/zoom gestures and provides filtering options through the sidebar.

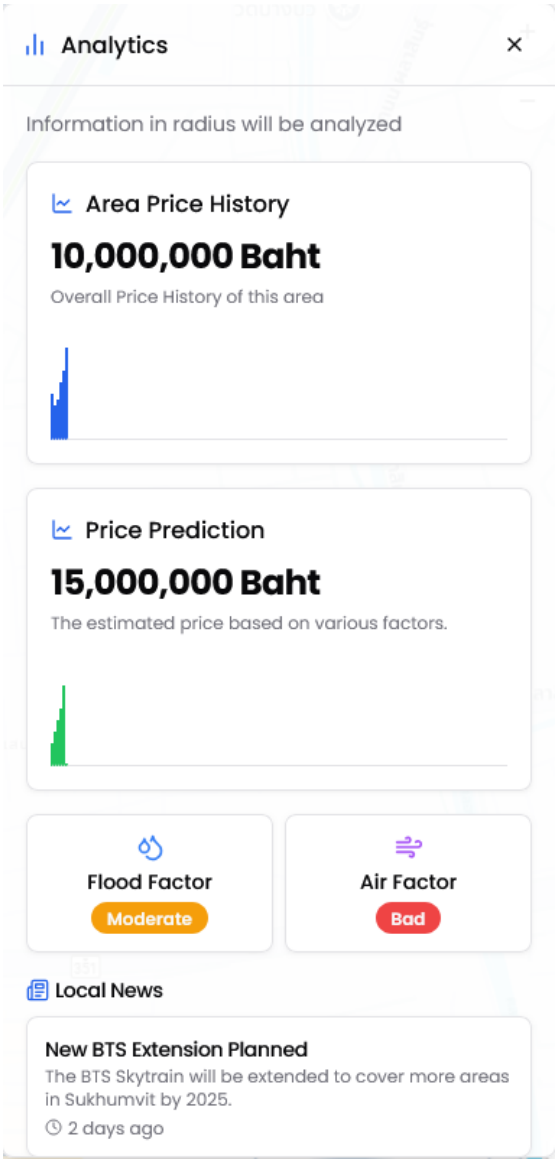

Figure 3.16: Map Analytics Overlay

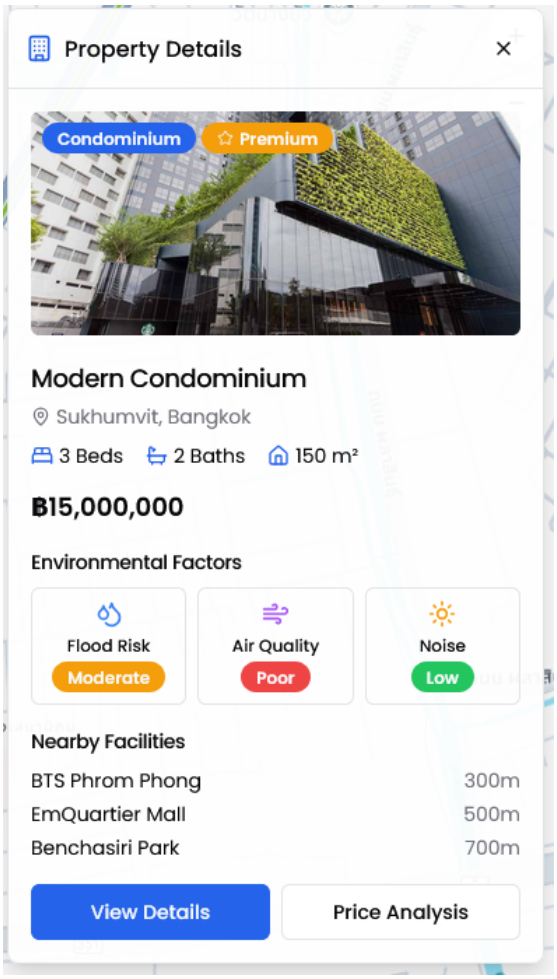

Figure 3.17: Property Detail Overlay

Figure 3.16 shows environmental factor visualization through heat maps and data layers. This overlay helps users analyze contextual factors like pollution levels directly on the map.

Figure 3.17 displays the popup that appears when users click map markers. This overlay provides immediate property information without requiring navigation away from the map.

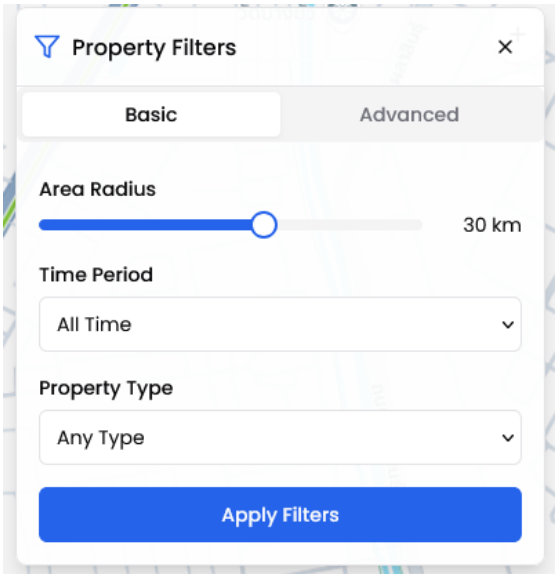

Figure 3.18: Property Filter Panel

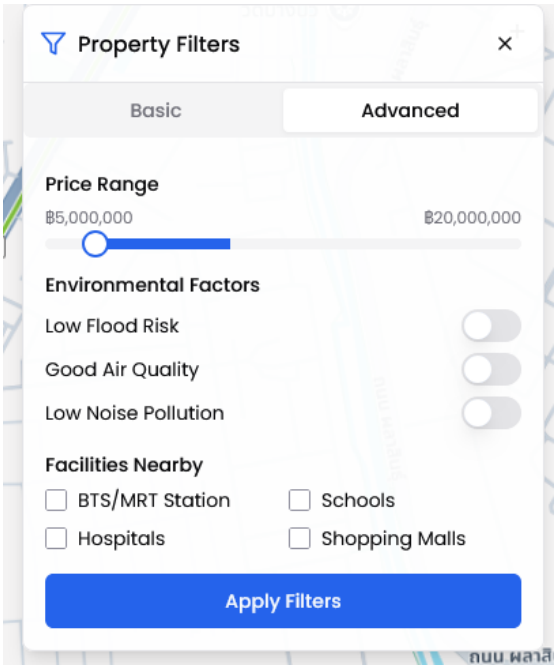

Figure 3.19: Advanced Property Filter Options

Figure 3.18 shows basic property filtering controls for price range, type and size parameters. The panel uses sliders and checkboxes for intuitive refinement of map results.

Figure 3.19 displays extended filtering options for specific amenities and neighborhood characteristics. These advanced parameters enable precise property matching based on detailed criteria.

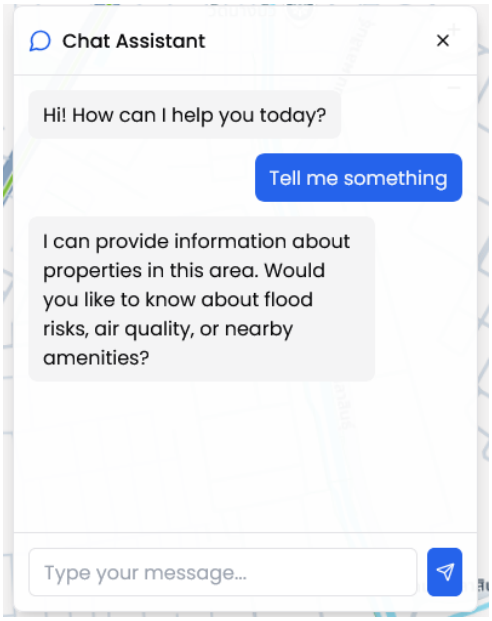

Figure 3.20: Interactive Chatbot Assistant

Figure 3.20 shows the conversational assistant for property search and analysis. The chatbot handles natural language queries about properties and market conditions.

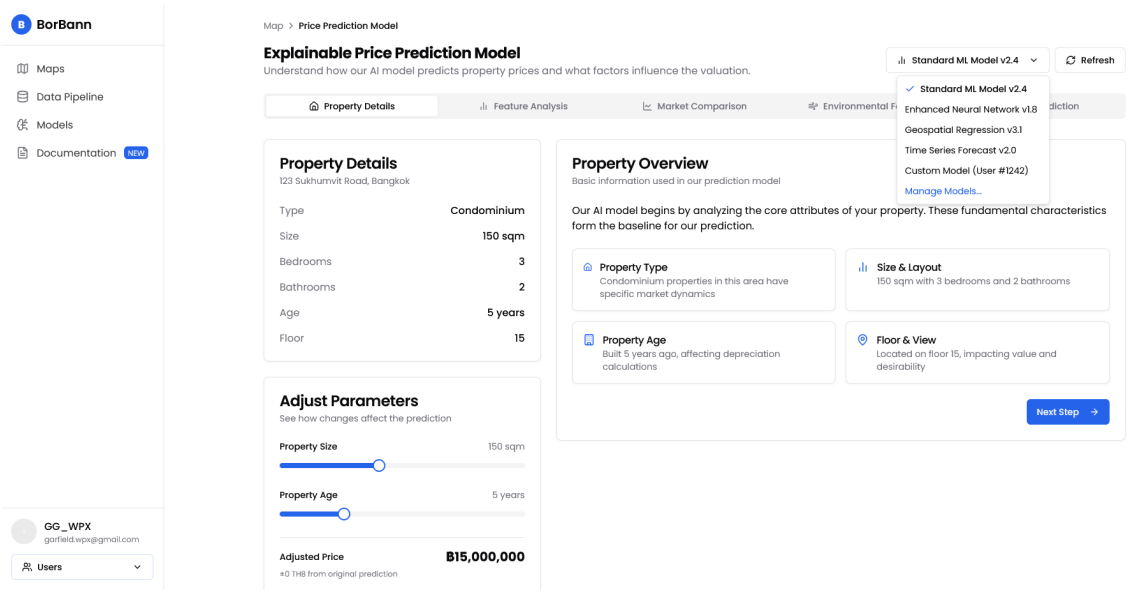

Figure 3.21: Explainable Price Prediction - Property Overview

Figure 3.21 displays the initial analysis of property attributes in the prediction model. The interface includes interactive parameter sliders that show how changes to property characteristics affect the predicted price.

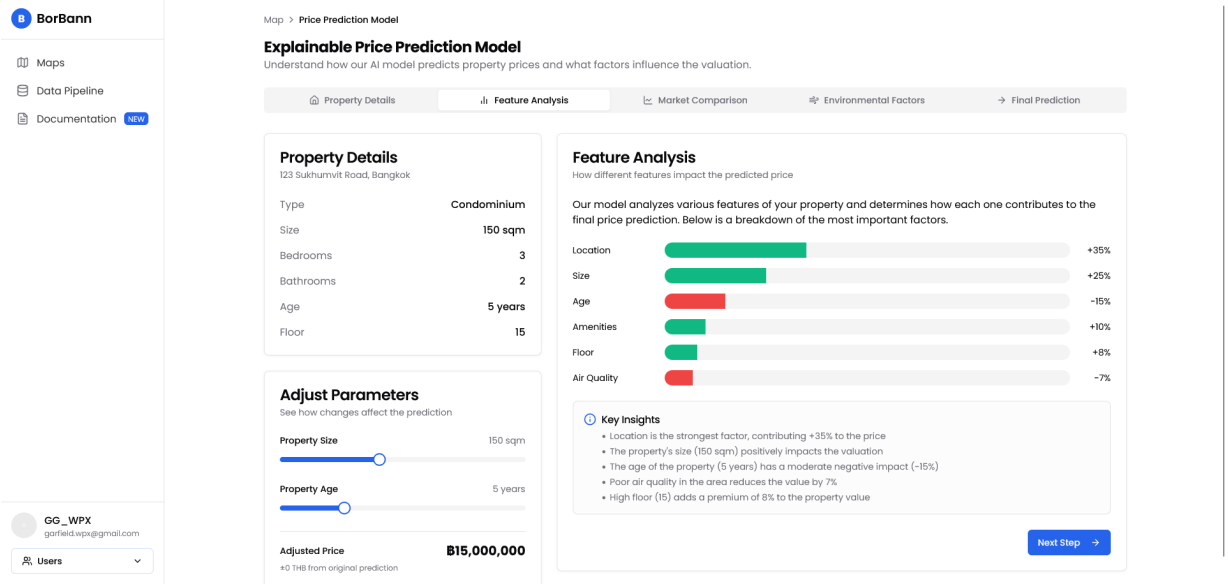

Figure 3.22: Explainable Price Prediction - Feature Importance Visualization

Figure 3.22 shows the contribution of different property features to the predicted price. The visualization uses color-coded bars to differentiate positive and negative factors with percentage impact values.

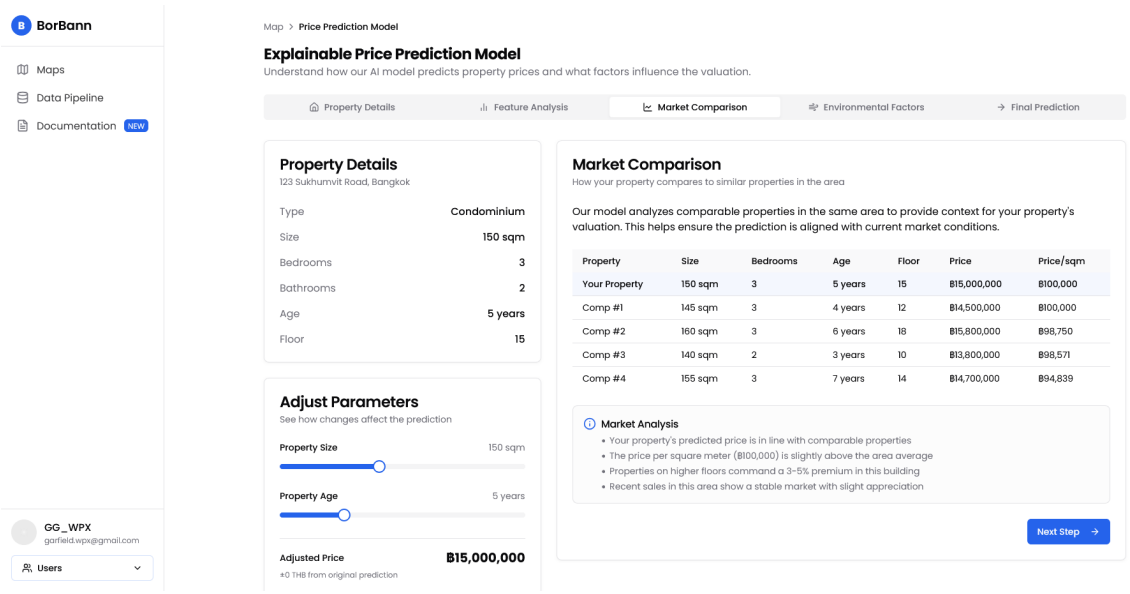

Figure 3.23: Explainable Price Prediction - Comparative Market Analysis

Figure 3.23 presents a comparison of the subject property against similar properties in the area. The table highlights key attributes and pricing metrics with supporting analysis of market positioning.

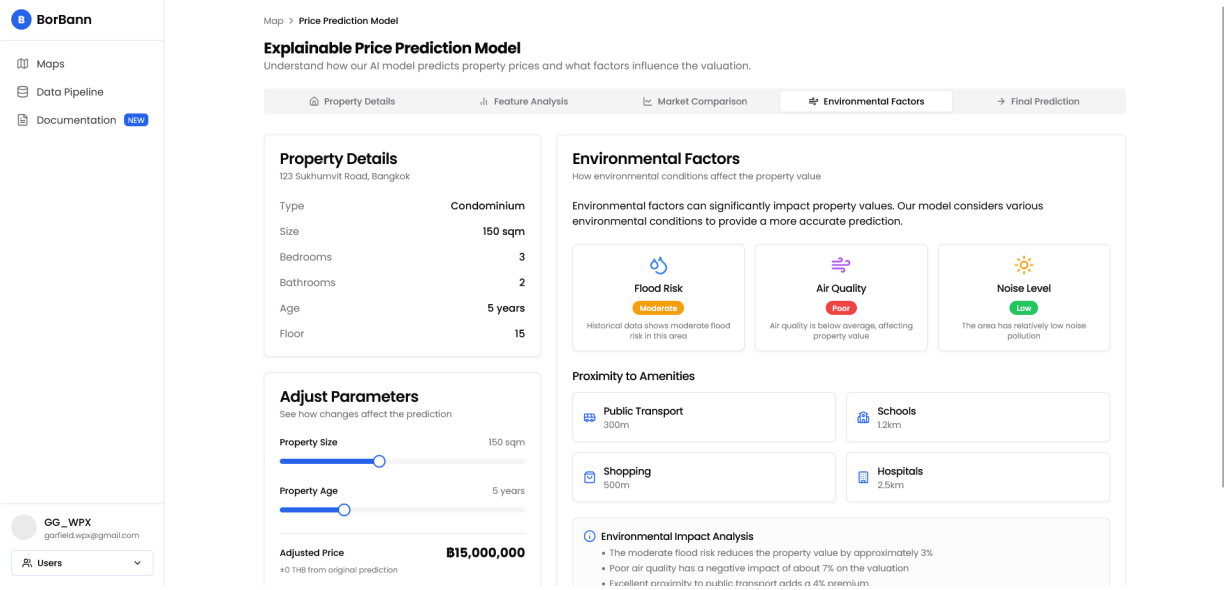

Figure 3.24: Explainable Price Prediction - Environmental Impact Analysis

Figure 3.24 analyzes environmental conditions and nearby amenities affecting property value. The interface displays flood risk, air quality, and proximity to key facilities with quantified impact on valuation.

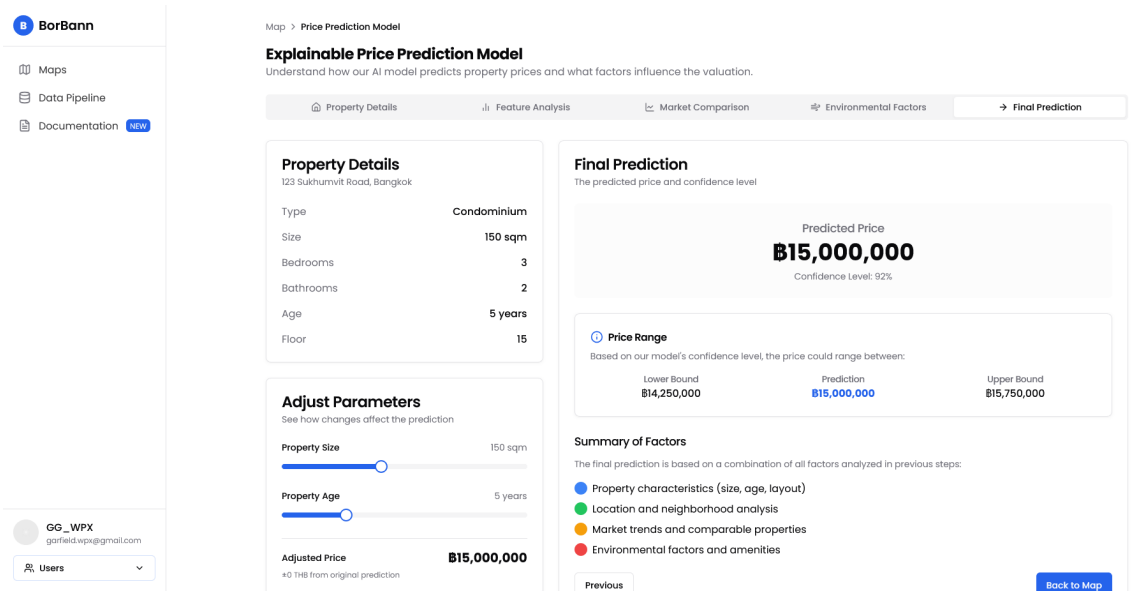

Figure 3.25: Explainable Price Prediction - Final Valuation with Confidence Range

Figure 3.25 shows the final price prediction with confidence metrics and range indicators. The summary section explains how the valuation synthesizes insights from all analytical components.

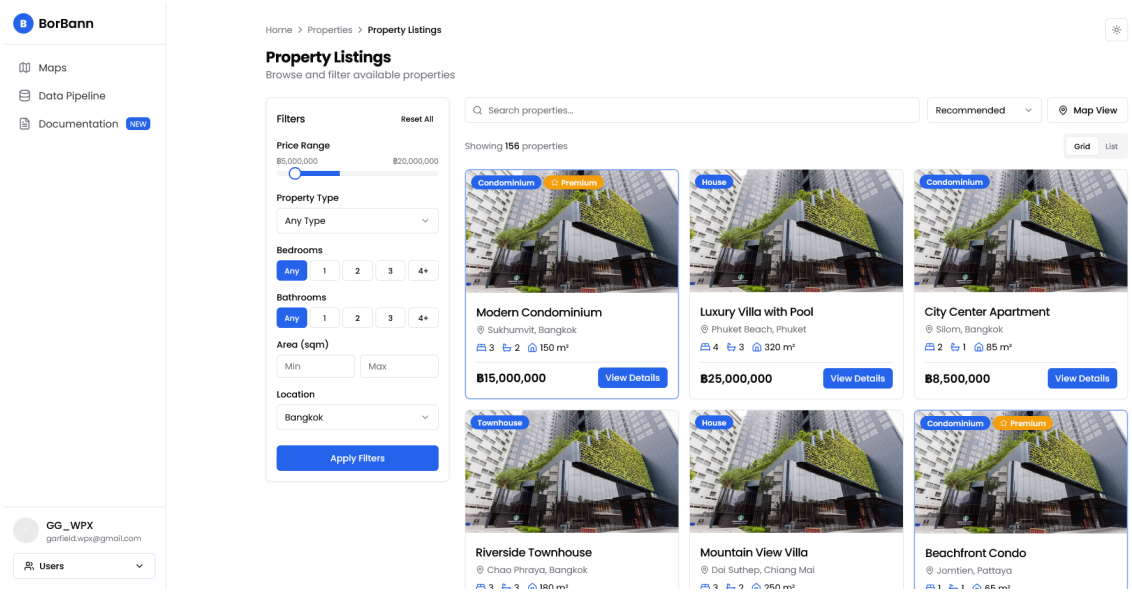

Figure 3.26: Property Listings Page

Figure 3.26 display property listings page with all listing that follow the filter tab the right.

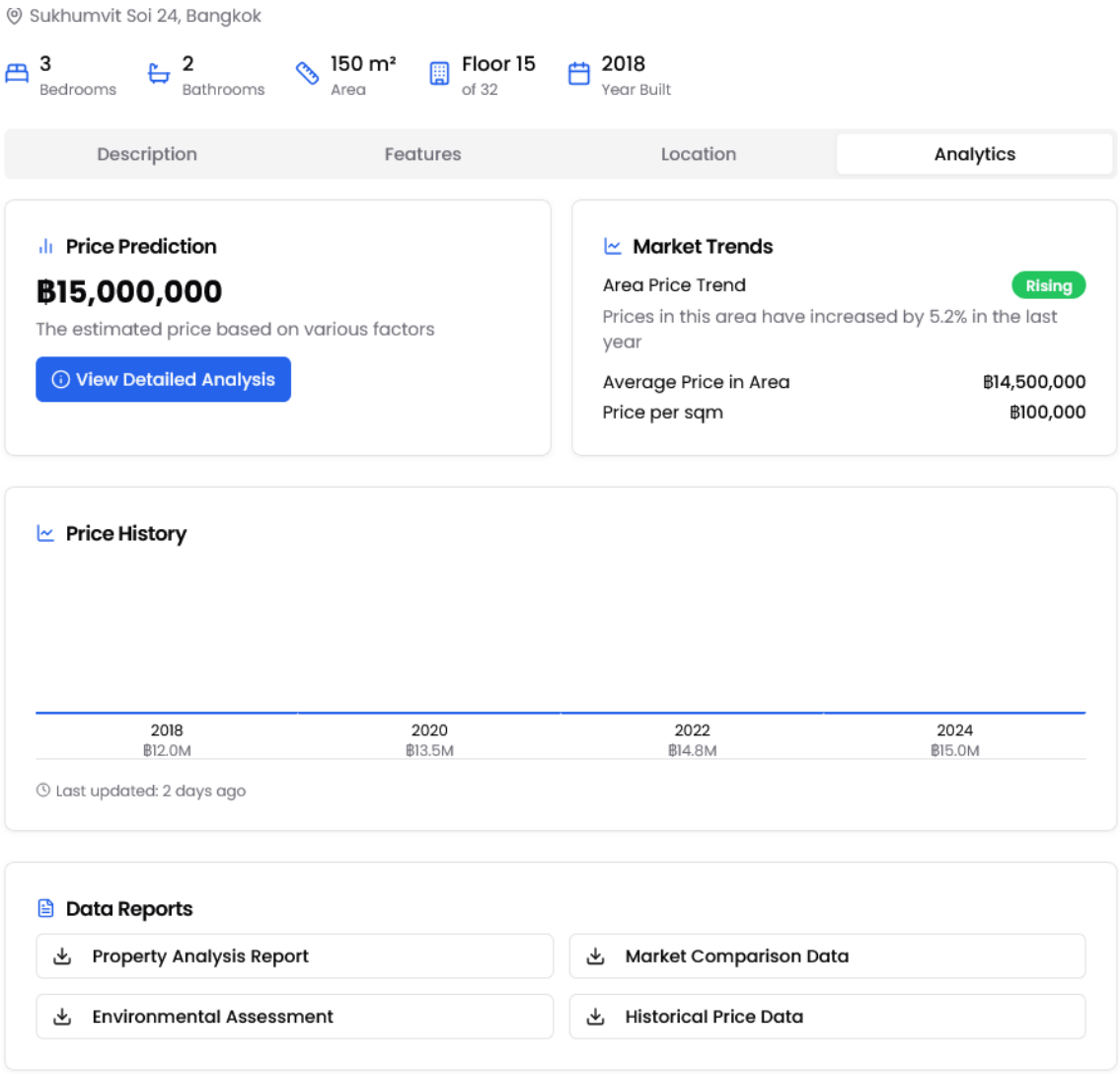

Figure 3.27: Property Listings Page - Analytic Tab

Figure 3.27 shows the analytic tab that show local context analytic of that specific listing.

Chapter 4

Software Architecture Design

4.1 Sequence Diagram

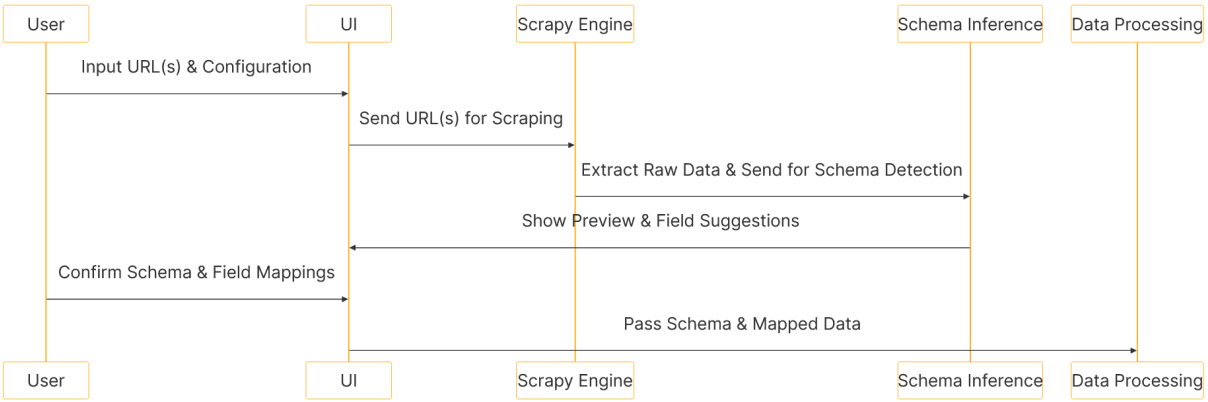

Figure 4.1: Sequence Diagram of Customizable Automated Data Integration Pipeline

The sequence diagram show the flow of interactions between key components of the customizable automated data integration pipeline. The process begins with the **User**, who inputs URLs and configuration details via the **UI**. The **UI** sends the provided URLs to the **Scrapy Engine** for scraping.

Once the data is extracted, the **Scrapy Engine** forwards the raw data to the **Schema Inference** module for automatic schema detection. The **Schema Inference** module analyzes the data and returns a preview, along with field mapping suggestions, to the **UI**. The **User** then reviews the suggested schema and field mappings in the **UI**.

Upon confirmation by the **User**, the **UI** passes the finalized schema and mapped data to the **Data Processing** pipeline, where further processing and integration of the collected data occur.

4.2 AI Component

The BorBann platform has AI components that enhance its analytical capabilities and user experience. These components work together to provide build a real estate data platform adapt to the Thai market context. Each component integrate with each other to form a comprehensive analytics system:

- The Data Integration Pipeline feeds data to the Local Contextual Analytics system
- Local Contextual Analytics provides features to the Explainable Price Prediction Model
- The Retraining Model uses data from the pipeline to create customized predictions
- All components share a common data model that enables smooth information flow

Explainable Price Prediction Model

The price prediction model delivers property price predictions with clear explanations of contributing factors.

Model Architecture

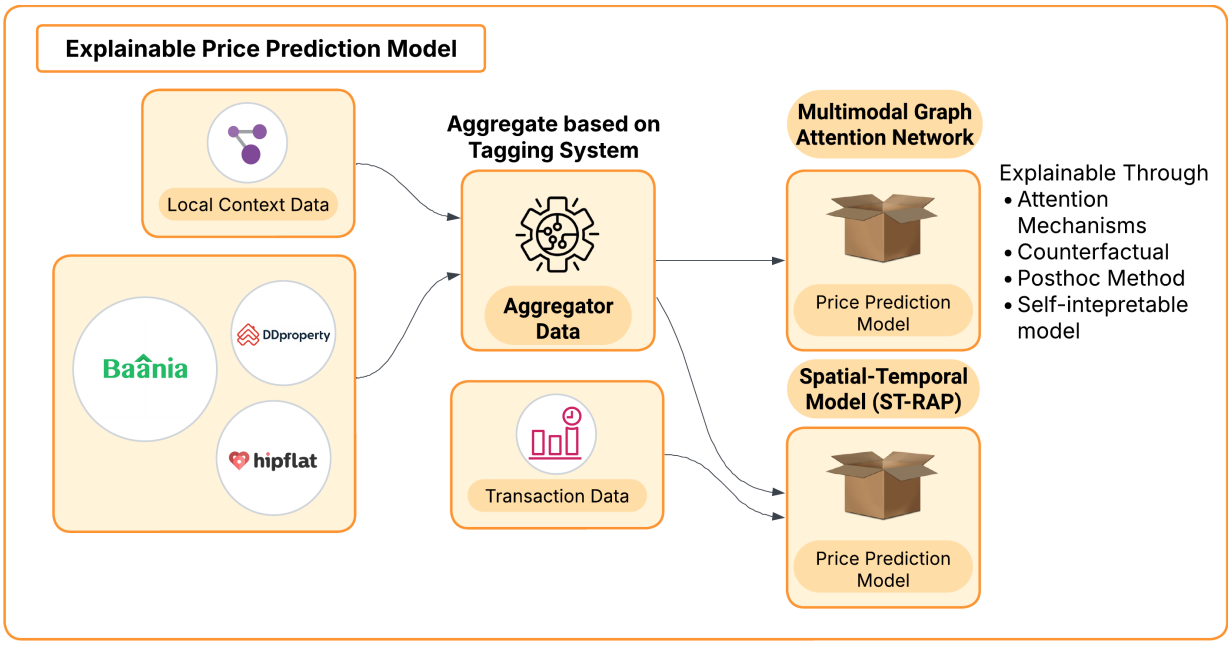

Figure 4.2: Explainable Price Prediction Data Flow

Figure 4.2 illustrates the architecture of the Explainable Price Prediction Model. The diagram shows how data flows from multiple sources (Baania, DDproperty, hipflat) and local context data into an Aggregator component that organizes information based on a tagging system. The aggregated data then feeds into two distinct prediction models:

a Multimodal Graph Attention Network and a Spatial-Temporal Model (ST-RAP). The explainability features are highlighted on the right, showing how the model provides transparency through attention mechanisms, counterfactual analysis, post-hoc methods, and self-interpretable modeling approaches.

- **Dual Modeling Approach:**
  - *Tree-based Model:* With tree-based gradient boosting models (XGBoost[4]).
  - *Multimodal Graph Attention Network:* With Multimodal Graph Attention Network[5] that is multi-modal model which can handle large graph.
  - *Spatial-Temporal Model:* Using ST-RAP[6] for facilities-property graph input and time series of real estate pricing data

Explainability Framework

- **In-model Attention:** With attention mechanism, we can look into weighted contributions of nodes, edges, or modalities
- **Post-hoc Explanation Module:** Uses KernelSHAP[7] (Algorithm to approximate SHapley Additive exPlanations) value calculation for feature attribution.
- **Self-Interpretable Components:** Extracts rules from complex models and implements decision tree surrogate models that approximate complex model behavior

User Interface Components

- **Interactive Visualizations:** Feature importance charts, and price trend projections with confidence intervals
- **Explanation Generation:** Natural language generation of price explanations and highlighting of key value drivers

Local Contextual Analytics

This component analyzes environmental conditions and proximity factors to evaluate property context and risk:

Model Architecture

Figure 4.3 depicts the Local Contextual Analytics system. The diagram shows various data sources including GISTDA, Bangkok Metropolitan Administration, news outlets, and property listing platforms (Baania, hipflat) feeding into an Aggregator Service. This service collects different types of data through web scraping, API fetching, and dataset downloads. The collected data undergoes preprocessing with outlier detection, value imputation, and feature engineering before being stored in specialized databases (Vector, Time-Series,

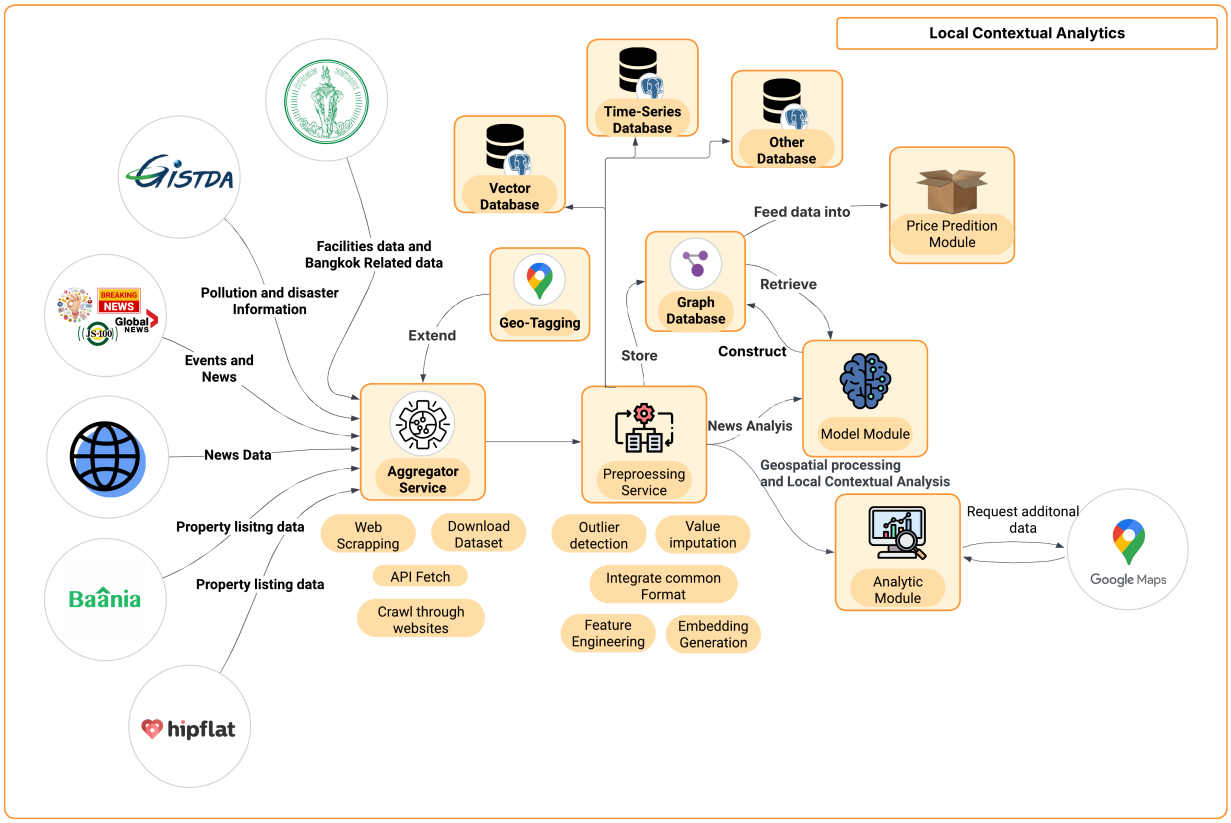

Figure 4.3: Local Contextual Analytic Data Flow

Graph). The Model Module and Analytic Module process this information to provide geospatial analysis and local contextual insights that ultimately feed into the Price Prediction Module.

- **Graph Database:** Stores heterogeneous data in Neo4J or other knowledge graph database
- **Vector Database:** Stores embedding from preprocessing unit within PostGreSQL with pgvector that will provide vector database capabilities
- **Timeseries Database:** Stores time-series data withtin Timescale
- **Relational Database:** Stores relational data in PostgreSQL
- **NoSQL Database:** Stores unstructured data within MongoDB
- **Model Module:**
  - Heterogeneous Graph Construction using k-NN or GCN
  - GCN models for spatial relationships
  - NLP models for news analysis
- **Price Prediction Module:** Integrated model consuming all processed features

Analytics Capabilities

- **Climate and Environmental Analysis:** Evaluates climate risk, pollution levels, and disaster assessment
- **Proximity Analysis:** Evaluates nearby locations and facilities, analyzing relationships between facilities
- **News Integration:** Incorporates local news into property assessment

Customizable Automated Data Integration Pipeline

This pipeline enables non-technical users to connect any data source into a unified system:

LLM Module

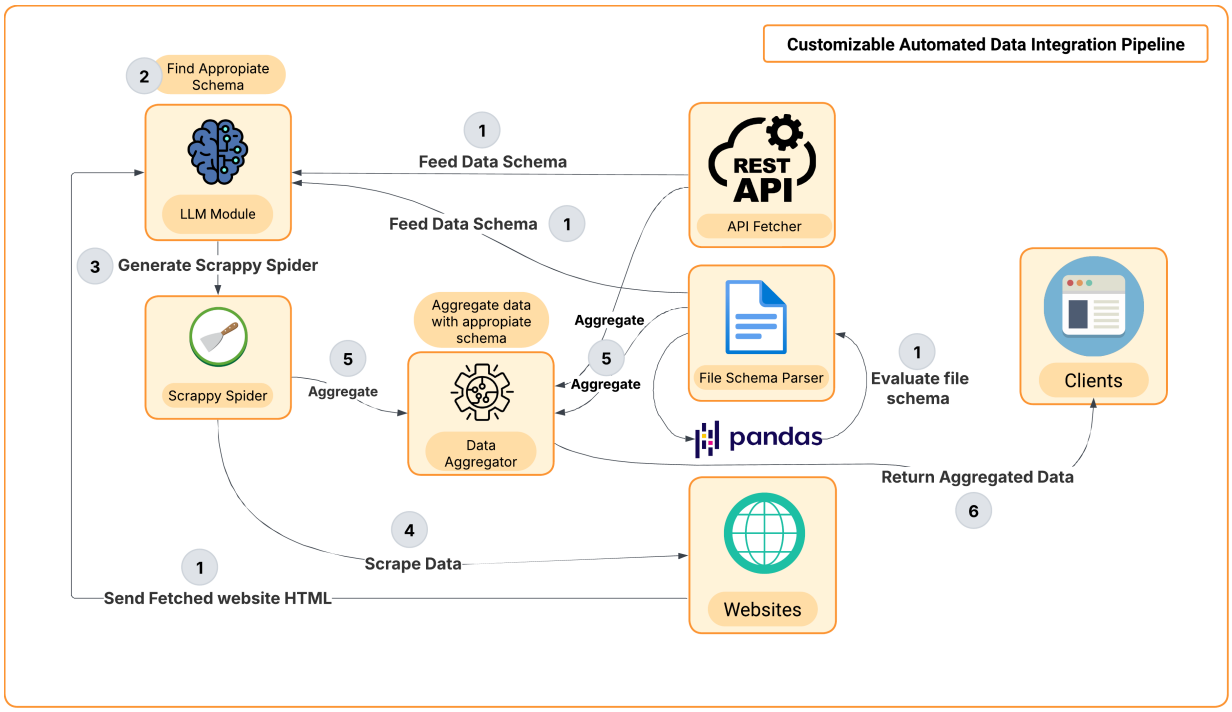

Figure 4.4: Customizable Automated Data Pipeline Data Flow

Figure 4.4 presents the Customizable Automated Data Integration Pipeline. The workflow begins with data sources (API Fetcher, File Schema Parser, Websites) being evaluated. An LLM Module analyzes these sources to find appropriate schemas (step 2), then generates Scrappy spiders for web scraping (step 3). The spiders extract data from websites (step 4), while the Data Aggregator combines information from all sources (step 5). Finally, the aggregated data is returned to clients (step 6). This pipeline enables non-technical users to integrate diverse data sources through an automated, LLM process.

- Analyzes data sources to find appropriate schemas

- For websites, generates Scrappy spider configurations automatically
- Bridges the gap between unstructured and structured data

Data Processing

- **Scrappy Spider:** Dynamically generated web scrapers based on LLM analysis that extract targeted data from websites according to the determined schema
- **Data Aggregator:** Central component that combines data from all sources, harmonizes different schemas into a unified structure, and uses appropriate schema mapping determined by the LLM

Retraining Model with Pipeline Data

This component allows users to create custom prediction models by combining their pipeline data with platform data:

Technical Implementation

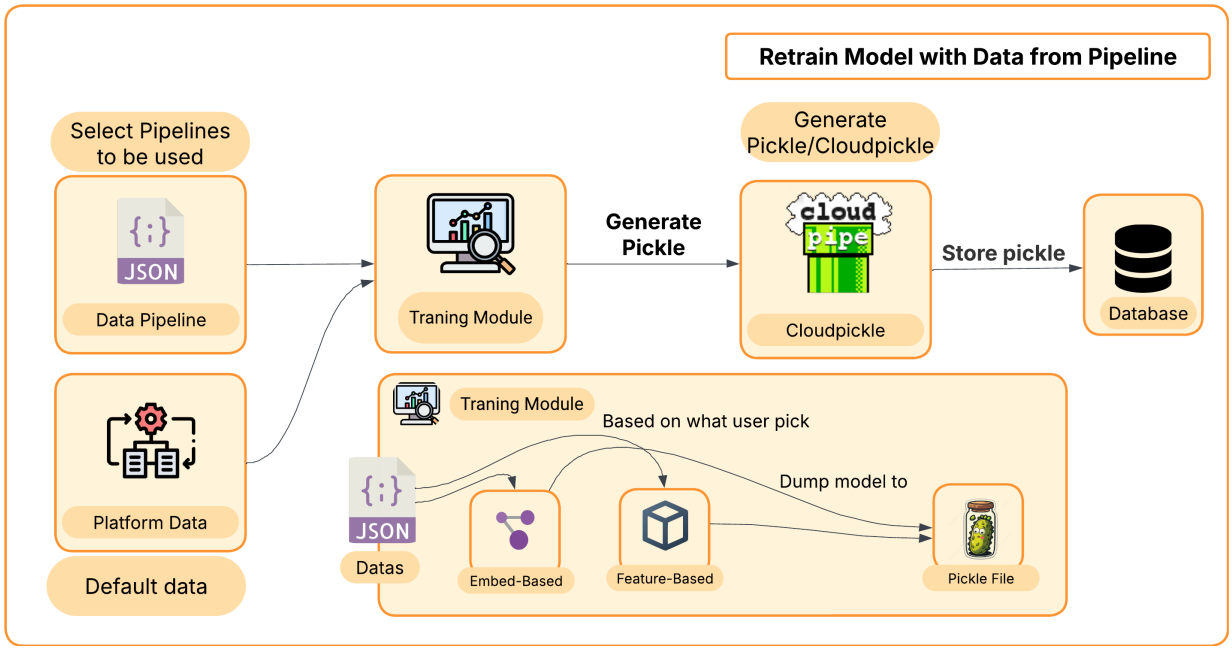

Figure 4.5: Retrain Model with Data from Pipeline Data Flow

Figure 4.5 shows the Model Retraining Pipeline. The diagram illustrates how users can select pipelines and combine them with platform data to train custom models. The Training Module processes this data and allows users to choose between embed-based or feature-based modeling approaches. The system generates pickle/cloudpickle files of the trained models, which are then stored in a database for future use. This component empowers users to create specialized prediction models tailored to their specific data and requirements.

- **Training Engine:** Implements automated training engine that rely on embedding-based models to avoid problem with unmatch input feature and tree-based models (for structured data), and builds parameter tuning system using random search if user doesn't specify ones.
- **Model Serialization:** Creates versioned model files for persistence and stores serialized models in pickle format in database with metadata
- **Deployment Framework:** Provides APIs for model inference, enables batch prediction, and implements monitoring for model drift detection

References

[1] J. Research, "Global real estate perspective february 2025: A positive but nuanced outlook for 2025," *JLL Global Real Estate Perspective*, February 2025. [Online]. Available: https://www.jll.co.th/en/trends-and-insights/research/global/gmp

[2] Nationthailand, "8 key trends shaping thai real estate after rollercoaster year," *Nationthailand*, January 2025. [Online]. Available: https://www.nationthailand.com/business/property/40044774

[3] B. Post, "Real estate information center targets better data gathering to head off crises," *Bangkok Post*, Jule 2024. [Online]. Available: https://www.bangkokpost.com/property/2836667/real-estate-information-center-targets-better-data-gathering-to-head-off-crises

[4] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. ACM, Aug. 2016, p. 785–794. [Online]. Available: http://dx.doi.org/10.1145/2939672.2939785

[5] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2018. [Online]. Available: https://arxiv.org/abs/1710.10903

[6] H. Lee, H. Jeong, B. Lee, K. Lee, and J. Choo, "St-rap: A spatio-temporal framework for real estate appraisal," 2023. [Online]. Available: https://arxiv.org/abs/2308.10609

[7] M. Mayer and D. Watson, *kernelshap: Kernel SHAP*, 2024, r package version 0.7.1. [Online]. Available: https://github.com/ModelOriented/kernelshap