PRODUCT REQUIREMENTS DOCUMENT

Navi Mumbai House Price Predictor

An AI-Powered Residential Property Valuation Platform

Version 1.0 | February 2026
Status: Draft for Review
Classification: Internal / Academic

1. Product Title & Version

Field	Details
Product Name	Navi Mumbai House Price Predictor
Version	1.0 (Initial Release)
Document Type	Product Requirements Document (PRD)
Date	February 2026
Author(s)	Project Team
Status	Draft for Review
Review Cycle	Bi-weekly
Target Release	Q3 2026

2. Executive Summary

2.1 Purpose of the Product

The Navi Mumbai House Price Predictor is a machine learning-powered web application designed to provide accurate, data-driven residential property price estimates within the Navi Mumbai metropolitan region of Maharashtra, India. The application enables users to input property parameters and receive an instant, AI-generated valuation — removing the opacity that has historically characterized real estate pricing in India.

2.2 Problem Being Solved

Residential property pricing in Navi Mumbai is highly fragmented, inconsistent, and often influenced by broker speculation rather than empirical market data. Prospective buyers, sellers, and investors lack a reliable, accessible tool to benchmark property values against comparable listings. This results in significant information asymmetry, financial risk, and poor decision-making.

This application addresses the problem by leveraging historical transaction data and machine learning algorithms to generate objective price predictions based on measurable property attributes.

2.3 Target Users

- First-time homebuyers seeking fair price benchmarks
- Real estate agents looking to validate or substantiate listing prices
- Investors and developers evaluating property ROI potential
- Banking and financial institutions requiring quick property appraisals
- Academic researchers studying real estate market trends

3. Background / Problem Statement

3.1 Challenges in Estimating House Prices

Accurate property valuation in urban India is a multi-dimensional challenge involving numerous interdependent variables. The following structural challenges make manual or intuition-based pricing unreliable:

- Rapid urbanization leads to frequent price volatility in micro-markets within the same city
- Lack of a centralized, publicly accessible database of verified transaction prices
- High dependency on real estate brokers who may present biased pricing
- Multiple property attributes — location, floor number, age, amenities, connectivity — interact non-linearly to determine price
- Price variation between adjacent localities can be as high as 30–40%, making general heuristics ineffective

3.2 Why Navi Mumbai-Specific Prediction is Valuable

Navi Mumbai is a planned satellite city developed by CIDCO (City and Industrial Development Corporation) that has emerged as one of the fastest-growing real estate markets in Maharashtra. Several factors make a localized prediction tool particularly impactful:

- Navi Mumbai comprises over 25 distinct nodes (Kharghar, Vashi, Panvel, Belapur, etc.) each with unique price dynamics
- The upcoming Navi Mumbai International Airport and proposed metro corridors are creating speculative price surges in specific zones
- A dedicated model trained on Navi Mumbai data outperforms generalized Mumbai or pan-India models due to local variable specificity
- The region attracts a large volume of middle-income home seekers, who would benefit most from an unbiased, easy-to-use pricing tool

4. Objectives

4.1 Business Objectives

- Deliver a Minimum Viable Product (MVP) within 6 months with core price prediction functionality
- Achieve a prediction accuracy (R^2 score) of 0.85 or above on the holdout test dataset
- Provide a scalable architecture capable of incorporating new training data quarterly
- Establish credibility as a data-backed real estate tool among Navi Mumbai real estate stakeholders
- Generate a platform suitable for monetization through premium features, API licensing, or institutional partnerships

4.2 User Objectives

- Receive an instant, reasonably accurate price estimate without reliance on a broker
- Understand which features (BHK, location, amenities) most influence property pricing

- Compare estimated prices across different localities and configurations
- Make informed buy, sell, or invest decisions backed by data

5. Scope

5.1 In-Scope Features

- Price prediction engine using a trained ML regression model
- User-facing web interface with structured input form
- Display of predicted price, confidence range, and comparable property data
- Input validation and error messaging for user-entered data
- Basic analytics dashboard showing price trends by locality
- Model training pipeline using historical Navi Mumbai property listing data
- Backend REST API serving predictions to the frontend
- Responsive design for mobile and desktop browsers

5.2 Out-of-Scope Items

- Commercial real estate or industrial property valuation
- Properties outside Navi Mumbai (e.g., Thane, Mumbai Suburban, Pune)
- Real-time integration with live property listing portals
- User account creation, authentication, or saved searches (Phase 2)
- Legal due diligence, title verification, or mortgage advisory services
- Rental yield prediction (planned for future versions)

6. Target Users / Personas

Persona	Description	Goal	Pain Point
Home Buyer — Rohan, 32	IT professional, first-time buyer, budget of ₹60–80 L	Determine fair market price for a 2 BHK in Kharghar	Cannot trust broker quotes; overwhelmed by price variation
Real Estate Agent — Priya, 41	Experienced agent with 15+ years in Navi Mumbai market	Quickly validate client asking prices for listings	Manual comparisons are time-consuming and inconsistent
Property Investor — Vikram, 48	High-net-worth individual exploring Panvel for investment	Identify undervalued zones with high appreciation potential	Lacks analytical tools; relies on anecdotal market intelligence
Bank Loan Officer — Sneha, 36	Evaluates property loan applications at a leading NBFC	Conduct a quick preliminary property appraisal	Formal appraisals take 5–7 days; needs faster reference estimates

7. User Stories

Functional User Stories

ID	As a...	I want to...	So that...
US-01	Home Buyer	Enter property details (location, BHK, area, floor) and get a predicted price	I can evaluate whether a listing is fairly priced before negotiating
US-02	Real Estate Agent	Input client property parameters and see comparable price range	I can justify listing prices to sellers and buyers with data
US-03	Investor	View price trend charts by locality and BHK type	I can identify high-appreciation micro-markets in Navi Mumbai
US-04	All Users	Receive a confidence interval with the price estimate	I understand the uncertainty margin and can plan accordingly
US-05	All Users	See a breakdown of which features most affect the predicted price	I understand the drivers behind the valuation
US-06	All Users	Get clear validation messages if I enter invalid or incomplete data	I can correct my inputs quickly without confusion
US-07	Bank Officer	Download a prediction summary as a PDF report	I can attach a quick reference estimate to the loan file
US-08	Developer / Admin	Retrain the ML model with updated listing data via an admin panel	The prediction engine remains accurate as market conditions evolve

8. Functional Requirements

8.1 User Input Fields

The application shall provide a structured input form with the following mandatory and optional fields:

Field	Type	Options / Constraints	Required
Node / Locality	Dropdown	25+ Navi Mumbai nodes (Kharghar, Vashi, Panvel, etc.)	Yes
Property Type	Radio	Apartment / Independent House / Villa / Studio	Yes
BHK Configuration	Dropdown	1 BHK, 2 BHK, 3 BHK, 4+ BHK	Yes
Built-Up Area (sq. ft.)	Numeric Input	100 – 10,000 sq. ft.; integer only	Yes
Floor Number	Numeric Input	0 (Ground) to 50	Yes
Total Floors in Building	Numeric Input	1–60; must be >= Floor Number	Yes
Age of Property (years)	Dropdown	Under Construction, 0–5, 5–10, 10–20, 20+	Yes
Car Parking	Radio	None / 1 / 2+	Yes
Furnishing Status	Radio	Unfurnished / Semi-Furnished / Fully Furnished	Yes
Amenities	Multi-Checkbox	Gym, Swimming Pool, Clubhouse, 24/7 Security, Power Backup, Lift	No
Proximity to Metro (km)	Numeric Input	0.1–20 km	No
RERA Registered	Toggle	Yes / No	No

8.2 Price Prediction Engine

- The system shall use a trained regression model (e.g., Gradient Boosting or Random Forest) to generate price predictions based on validated user inputs
- The model shall output: (1) Predicted price in INR (₹), (2) Price per sq. ft., (3) 90% confidence interval (lower and upper bounds)
- Prediction latency shall not exceed 2 seconds under normal load conditions
- The model shall be versioned; each deployment shall log model version used for traceability

8.3 Result Display

- Predicted price shall be displayed prominently in large, formatted INR currency (e.g., ₹68,50,000 or ₹1.2 Cr)
- A horizontal bar or gauge shall visually represent the confidence range
- A feature importance chart (horizontal bar chart) shall display the top 5 factors influencing the prediction

- A locality price comparison table shall display average prices for the selected BHK type across nearby nodes

8.4 Input Validation

- All required fields must be filled before form submission; empty required fields trigger inline error messages
- Built-up area must be a positive integer within the acceptable range; decimal inputs shall be rejected
- Floor number must not exceed total floors in building; cross-field validation shall be enforced
- The system shall prevent form submission if validation fails and highlight erroneous fields in red

9. Non-Functional Requirements

Category	Requirement	Target Metric
Performance	Prediction response time from form submission to result display	< 2 seconds (p95)
Performance	Page load time on 4G mobile network	< 3 seconds
Scalability	Concurrent users supported without degradation	Up to 500 concurrent
Accuracy	Model R^2 Score on test dataset	>= 0.85
Accuracy	Mean Absolute Percentage Error (MAPE)	<= 12%
Security	HTTPS enforced across all endpoints	100% HTTPS
Security	User input sanitized to prevent injection attacks	OWASP Top 10 compliance
Usability	Task completion rate for first-time users (usability test)	>= 90%
Usability	System Usability Scale (SUS) Score	>= 75 (Good)
Availability	System uptime SLA	>= 99.5% monthly
Maintainability	Model retraining pipeline execution time	< 4 hours
Accessibility	WCAG 2.1 Level AA compliance	Full compliance

10. System Features

10.1 Prediction Model

The core ML engine shall be based on an ensemble regression approach. Model selection shall be determined through comparative evaluation of the following candidate algorithms:

- Linear Regression (baseline benchmark)
- Random Forest Regressor
- Gradient Boosting Regressor (XGBoost / LightGBM)
- Support Vector Regression (SVR) — optional benchmark

The final deployed model shall be selected based on R^2 score, MAPE, and cross-validation performance. Hyperparameter tuning shall be performed using Grid Search or Bayesian Optimization.

10.2 Data Handling

- All training data shall be stored in a structured database with version control
- The system shall log each prediction request with input parameters, predicted output, and timestamp for audit and model feedback purposes
- Personally identifiable information (PII) shall not be collected or stored as part of prediction requests
- A scheduled monthly data ingestion pipeline shall refresh the training dataset

10.3 UI Behavior

- The form shall use progressive disclosure — advanced fields (amenities, proximity, RERA) shall be collapsed under an 'Advanced Options' toggle
- Real-time field hints (e.g., average area for 2 BHK in Kharghar: 900–1100 sq. ft.) shall be displayed contextually
- Result panel shall animate into view post-submission with a brief loading skeleton to signal processing
- A 'Reset' button shall clear all fields and return the form to its default state

11. Data Requirements

11.1 Types of Data Needed

Data Type	Description	Format
Property Listings	Historical residential listing data with price, area, BHK, and location	CSV / JSON
Transaction Records	Registered sale deed prices from IGRS Maharashtra	PDF / Structured DB
Geospatial Data	Latitude/longitude of properties and landmarks (metro, hospitals, schools)	GeoJSON / CSV
Amenity Tags	Binary flags for listed amenities per property	CSV / JSON
Market Indices	Quarterly price index by locality (NHB Residex or equivalent)	CSV

11.2 Data Sources

- Publicly available property listings from portals such as 99acres, MagicBricks, and Housing.com (scraped or licensed)
- Maharashtra IGR (Inspector General of Registration) for verified transaction data
- CIDCO official portal for development zone and node classification data
- OpenStreetMap for geospatial proximity features
- NHB Residex for quarterly market index benchmarking

11.3 Data Preprocessing

- Removal of duplicate listings and outlier prices (> 3 standard deviations from local mean)
- Imputation of missing values using median/mode substitution or KNN imputation
- One-hot encoding for categorical variables (locality, furnishing status, property type)
- Feature engineering: price per sq. ft., floor ratio (floor / total floors), amenity score (weighted sum)
- Train/Validation/Test split of 70% / 15% / 15% with stratification by locality and BHK
- Min-max normalization applied to continuous features for model compatibility

12. Technology Stack (Suggested)

Layer	Technology	Justification
Frontend	React.js + Tailwind CSS	Component-based UI, rapid development, responsive design
Charts / Visualization	Recharts or Chart.js	Lightweight, declarative charting for React
Backend / API	Python FastAPI	High performance, async support, native ML library integration
ML Libraries	scikit-learn, XGBoost, LightGBM, pandas, NumPy	Industry-standard Python ML ecosystem
Model Serialization	joblib / pickle	Efficient serialization of trained sklearn / XGBoost models
Database	PostgreSQL	Reliable relational DB for training data and prediction logs
Data Processing	Apache Spark (optional) or pandas	Scalable data wrangling for large listing datasets
Model Tracking	MLflow	Experiment tracking, model versioning, and registry
Deployment	Docker + Nginx + AWS EC2 / GCP Cloud Run	Containerized, portable, cloud-agnostic deployment

Layer	Technology	Justification
CI/CD	GitHub Actions	Automated testing, linting, and deployment pipeline
Monitoring	Prometheus + Grafana (optional)	Real-time system performance and model drift monitoring

13. Success Metrics / KPIs

KPI	Definition	Target	Measurement Method
Model Accuracy (R^2)	Proportion of price variance explained by the model	>= 0.85	Evaluation on holdout test set
Mean Absolute % Error	Average prediction deviation from actual price	<= 12%	Evaluation on holdout test set
Response Time (p95)	95th percentile latency for prediction requests	< 2 seconds	Backend performance monitoring
Page Load Time	Time to interactive on mobile (4G)	< 3 seconds	Lighthouse / WebPageTest
Task Completion Rate	Users who successfully receive a prediction result	>= 90%	Usability testing sessions
SUS Score	System Usability Scale score (0–100)	>= 75	Post-session user surveys
Monthly Active Users	Unique users who submit at least one prediction per month	>= 500 by Month 3	Analytics platform
Data Freshness	Age of most recent training data used in production model	<= 90 days	MLflow model metadata

14. Risks & Mitigation

Risk	Likelihood	Impact	Mitigation Strategy
Insufficient or low-quality training data from Navi Mumbai market	Medium	High	Supplement scraped data with synthetic augmentation; establish data partnerships with portals

Risk	Likelihood	Impact	Mitigation Strategy
Model prediction accuracy below acceptable threshold (< 0.80 R^2)	Medium	High	Iterative model improvement; include ensemble stacking; expand feature set with geospatial data
Data licensing or scraping restrictions from listing portals	High	Medium	Prioritize IGRS public data; explore data licensing agreements; use legally available public datasets
Rapid market price changes reducing model validity	Medium	Medium	Implement quarterly retraining pipeline; add market index as a real-time feature
User mistrust of AI-generated valuations	Medium	Medium	Display confidence intervals; explain feature importance; add disclaimer about advisory nature
Scalability bottlenecks under high concurrent load	Low	Medium	Use containerized deployment with horizontal auto-scaling; conduct load testing pre-launch
Regulatory changes affecting real estate data availability	Low	High	Monitor regulatory environment; diversify data sources; maintain data backup archives

15. Future Enhancements

Phase 2 — User Personalization

- User registration, profile management, and saved searches
- Price alert notifications when predicted prices for a saved configuration change significantly
- Comparison tool: side-by-side prediction for up to 3 property configurations

Phase 3 — Advanced Analytics

- Rental yield predictor and buy-vs-rent comparison calculator
- Locality investment score based on appreciation potential, infrastructure development, and connectivity
- Time-series forecasting to predict future price trajectories (6-month and 12-month outlook)

Phase 4 — Platform Expansion

- Expansion to other Maharashtra cities: Pune, Thane, Nashik
- API marketplace allowing third-party integrations (banks, brokerages, portals)

- Mobile application (iOS and Android) with offline caching
- Virtual property tour integration via 360-degree image embedding

16. Conclusion

The Navi Mumbai House Price Predictor addresses a well-defined, high-value problem in one of India's most dynamic real estate markets. By combining curated local property data with modern machine learning techniques, the application is positioned to deliver meaningful, transparent, and accurate property valuations to a diverse user base.

This Product Requirements Document establishes a clear foundation for the development team, stakeholders, and reviewers. The scope, functional requirements, technical architecture, and success metrics defined herein provide sufficient guidance to begin sprint planning and iterative development.

The product's emphasis on data transparency, model explainability, and user-centric design distinguishes it from general-purpose real estate portals and positions it as a trusted analytical tool for the Navi Mumbai property ecosystem. With phased feature enhancements and a robust data pipeline, the platform is designed to scale in utility, reach, and business value over time.

— End of Document —