

PRODUCT REQUIREMENTS DOCUMENT

# Navi Mumbai House Price Prediction Platform

*PropSight NM — AI-Powered Real Estate Valuation*

Version	1.0
Date	February 2025
Status	Draft — For Review
Author	Product Team
Stakeholders	Engineering, Data Science, Business Dev, Legal

## 1. Executive Summary

PropSight NM is an AI-driven house price prediction platform purpose-built for the Navi Mumbai real estate market. By aggregating historical transaction data, micro-location intelligence, infrastructure developments, and macroeconomic signals, the platform delivers accurate, explainable property valuations for buyers, sellers, agents, and financial institutions.

Navi Mumbai — spanning nodes such as Vashi, Kharghar, Panvel, Nerul, Airoli, and Belapur — is one of India's fastest-growing planned urban agglomerations. The market lacks a dedicated, data-rich valuation tool, creating significant friction for all stakeholders. PropSight NM addresses this gap with an ML-powered engine, interactive dashboards, and a developer-friendly API.

## 2. Problem Statement

### 2.1 Market Pain Points

- Buyers overpay or under-negotiate due to opaque, agent-driven pricing.
- Sellers lack objective benchmarks, leading to prolonged listing periods.
- Banks and NBFCs rely on manual, subjective appraisals for home loans — slow and error-prone.
- Real estate agents have no standardised tool for quick comparative market analysis (CMA).
- Existing pan-India platforms (99acres, MagicBricks) provide only aggregated data — no Navi Mumbai-specific depth.

### 2.2 Opportunity

Navi Mumbai saw over 42,000 residential registrations in 2023–24 (MahaRERA data), with a CAGR of ~11% in average property values over the last 5 years. Significant catalysts — Navi Mumbai International Airport (NMIA), Navi Mumbai Metro Line 1, and Mumbai Trans-Harbour Link (MTHL) — are actively reshaping micro-market valuations, creating an urgent need for a dynamic, forward-looking prediction engine.

### 3. Goals & Objectives

#### 3.1 Product Goals

- 1. Deliver price predictions with  $\leq 8\%$  Mean Absolute Percentage Error (MAPE) for residential properties.
- 2. Provide instant valuations (<3 seconds) for any residential unit in Navi Mumbai.
- 3. Support 10,000+ daily active users (DAU) within 12 months of launch.
- 4. Achieve a 4.2/5 user satisfaction score within 6 months of launch.

#### 3.2 Business Goals

- Generate B2B revenue via API subscriptions for banks, NBFCs, and PropTech platforms.
- Establish PropSight NM as the go-to valuation authority for Navi Mumbai within 18 months.
- Partner with MahaRERA and IGR Maharashtra for official data integration.

### 4. Target Users & Personas

Persona	Profile	Primary Need	Key Feature
Home Buyer	30–45 yr, first/second purchase	Fair price benchmarking	Instant Estimate + Comparables
Property Seller	Individual / Developer	Optimal listing price	Price Trend Dashboard
Real Estate Agent	RERA-registered broker	Quick CMA reports	Bulk Valuation + PDF Export
Bank / NBFC	Home loan underwriter	Automated appraisal	API Integration + Confidence Score
Investor / Fund	HNI / PE real estate fund	ROI & appreciation forecast	Heat Maps + Scenario Modelling

### 5. Feature Requirements

#### 5.1 Core Features (MVP — Phase 1)

##### F1: Instant Price Estimator

- User inputs: location (node/sector/plot), property type (1BHK/2BHK/3BHK/4BHK, flat/villa/row house), carpet area (sq ft), floor, age of building, parking, and furnishing status.
- System returns: predicted price range (low–mid–high), price per sq ft, confidence score (0–100%), and three comparable transactions.
- Prediction latency target: <3 seconds (P95).

## F2: Price Trend Dashboard

- Historical price trends by node (Vashi, Kharghar, Panvel, Nerul, Airoli, Belapur, Ulwe, Taloja, etc.) over 1M / 3M / 1Y / 3Y / 5Y periods.
- Price index visualised as interactive line charts (price per sq ft over time).
- Overlay of key macro events: MTHL inauguration, Metro Line 1 opening, NMIA construction milestones.

## F3: Comparable Sales (Comps) Engine

- Retrieve up to 10 recent comparable registered transactions within 1 km radius and  $\pm 15\%$  size variance.
- Data sourced from IGR Maharashtra (stamp duty registrations).
- Display on interactive map (Mapbox / Google Maps SDK).

## F4: Neighbourhood Scorecard

- Composite score (0–100) for: connectivity, schools, hospitals, retail, green spaces, crime index, and infrastructure pipeline.
- Sourced from CIDCO master plan, OpenStreetMap POI data, and Navi Mumbai Municipal Corporation (NMMC) datasets.

## 5.2 Advanced Features (Phase 2 — Q3 2025)

### F5: Future Price Forecasting

- 6-month and 12-month price forecast with confidence bands.
- Scenario modelling: impact of NMIA opening (expected 2025), Metro Phase 2, and new CIDCO residential schemes.

### F6: Investment Heat Map

- Colour-coded choropleth map showing expected appreciation rates across Navi Mumbai nodes.
- Filter by: budget range, property type, expected ROI, rental yield.

### F7: Bulk Valuation API

- RESTful API for B2B clients (banks, aggregators).
- Accepts CSV/JSON batch requests up to 500 properties per call.
- Rate limits: 1,000 API calls/day (standard), 10,000/day (enterprise).

- SLA: 99.9% uptime, <5s P99 latency.

F8: Valuation Report (PDF)

- Auto-generated, branded PDF report for agents and banks.
- Includes: estimated value, confidence score, comps table, trend chart, neighbourhood scorecard, and disclaimer.

6. Data Sources & ML Architecture

6.1 Data Sources

Data Source	Data Points	Refresh Frequency
IGR Maharashtra	Registered transaction price, date, area, property type	Daily
MahaRERA	Project approvals, developer track record, possession dates	Weekly
CIDCO / NMMC	Infrastructure pipeline, land use zones, approvals	Monthly
OpenStreetMap + Google Maps	POIs, school/hospital proximity, transit access	Monthly
NoBroker / 99acres / MagicBricks	Listing prices (ask prices) for demand signals	Daily
RBI / NHB	Home loan rates, housing index (NHB RESIDEX)	Monthly
Census / MOSPI	Demographic data, household income estimates	Annual

6.2 ML Model Architecture

- Primary Model: Gradient Boosting (XGBoost / LightGBM) trained on 5+ years of registered transactions (~180,000 records).
- Features: 40+ engineered features including location embeddings (geo-coordinates), temporal features, property attributes, and macro signals.
- Ensemble approach: Blend of GBM, Random Forest, and Neural Network (tabular) for improved robustness.
- Explainability: SHAP values surfaced in UI to show top 5 price drivers for each prediction.
- Retraining Cadence: Weekly retraining on incremental IGR data; full retraining monthly.
- Model Registry: MLflow on AWS SageMaker; A/B tested before production deployment.

7. Technical Architecture

7.1 System Components

- Frontend: React.js (Web) + React Native (iOS & Android); Mapbox for geospatial visualisations.
- Backend: FastAPI (Python) microservices; Node.js BFF (Backend for Frontend).
- ML Serving: AWS SageMaker endpoints with auto-scaling; Redis caching for frequent queries.
- Data Pipeline: Apache Airflow orchestration; dbt for data transformations; Snowflake as data warehouse.
- Infrastructure: AWS (Mumbai region ap-south-1); Terraform IaC; GitHub Actions CI/CD.
- Observability: Datadog APM + logging; PagerDuty alerting; Grafana dashboards.

7.2 Security & Compliance

- Data residency: All data stored in AWS ap-south-1 (Mumbai) in compliance with India's DPDP Act 2023.
- PII handling: No collection of user PII beyond email and mobile (optional); anonymised usage analytics.
- API security: OAuth 2.0 + API key authentication; rate limiting via AWS WAF; TLS 1.3 in transit.
- IGR data usage: Compliant with IGR Maharashtra data licensing terms; no redistribution of raw records.

8. Success Metrics & KPIs

Metric Category	KPI	Target	Timeline
Model Accuracy	MAPE on holdout test set	≤8%	At launch
Model Accuracy	Coverage (% properties predictable)	>90% of Navi Mumbai	At launch
Product Engagement	Daily Active Users	10,000	Month 12
Product Engagement	Estimates generated per day	25,000	Month 12
Product Quality	User satisfaction score (CSAT)	≥4.2 / 5	Month 6
Product Quality	P95 prediction latency	<3 seconds	At launch
Business	B2B API clients	20	Month 12
Business	Monthly Recurring Revenue	₹50L MRR	Month 18

9. Product Roadmap

Phase	Timeline	Deliverables	Go/No-Go Criteria
Phase 0 — Foundation	Month 1–2	Data pipeline setup, IGR/MahaRERA ingestion, baseline ML model (MAPE <12%), internal tooling	MAPE <12% on validation set
Phase 1 — MVP	Month 3–5	F1 Estimator, F2 Trends, F3 Comps, F4 Scorecard; Web app + Android beta	MAPE <8%, 500 beta users
Phase 2 — Growth	Month 6–9	F5 Forecast, F6 Heat Map, F8 PDF Reports, iOS app, RERA partnership	10K DAU, CSAT >4.0
Phase 3 — Monetisation	Month 10–15	F7 Bulk API, B2B onboarding, bank integrations, premium subscription tiers	₹20L MRR, 5 B2B clients

10. Risks & Mitigations

Risk	Severity	Impact	Mitigation
IGR data access disruption	High	Model staleness, loss of comparables	Cache 6 months of data; explore secondary data partnerships
NMIA/Metro delay shifts micro-market dynamics unexpectedly	Medium	Model drift, inaccurate forecasts	Monthly model retraining; add infrastructure delay scenarios to forecast model
Regulatory changes to data sharing (DPDP Act enforcement)	Medium	Data pipeline redesign required	Privacy-by-design from day 1; legal review of all data sources
Competition from funded PropTech startups entering Navi Mumbai	Medium	User acquisition cost increase	First-mover advantage; deepen CIDCO/MahaRERA partnerships; focus on B2B moat
Model accuracy gap in low-transaction micro-markets (e.g., Taloja)	Low	Poor UX in sparse zones	Hierarchical modelling; clearly communicate confidence intervals to users

11. Out of Scope

- Commercial / industrial property valuations (Phase 4+).
- Properties outside Navi Mumbai — Greater Mumbai, Thane, Pune excluded from MVP.
- Rental price prediction (separate product stream).
- Legal due diligence or title search (refer to legal partners).
- Real-time transaction facilitation or brokerage.

12. Appendix

12.1 Glossary

- MAPE: Mean Absolute Percentage Error — primary ML accuracy metric.
- IGR: Inspector General of Registration, Maharashtra — official property registration authority.

- MahaRERA: Maharashtra Real Estate Regulatory Authority.
- CIDCO: City and Industrial Development Corporation — Navi Mumbai's planning authority.
- NMIA: Navi Mumbai International Airport.
- MTHL: Mumbai Trans-Harbour Link (Atal Setu) — 21.8 km sea bridge connecting Mumbai and Navi Mumbai.
- CMA: Comparative Market Analysis.
- SHAP: SHapley Additive exPlanations — ML explainability framework.

## 12.2 Regulatory References

---

- Maharashtra Real Estate (Regulation and Development) Act, 2016.
- Digital Personal Data Protection Act, 2023 (India).
- IGR Maharashtra data licensing terms (2023 revision).
- NHB RESIDEX: National Housing Bank Residential Price Index.