Hotel Booking Cancellation Prediction System — Product Requirements Document

Date: June 5, 2026 **Author:** Sankar Kumar Palaniappan **Program:** MIT — No Code AI and Machine Learning (Great Learning) **Status:** Final **Version:** 2.0 (All evaluation gaps addressed — 16/16)

---

1. Executive Summary

Problem Statement

Hotels and online booking platforms lose significant revenue to last-minute cancellations — with industry cancellation rates at 33% and revenue loss reaching up to 15% of net revenue — yet have no reliable mechanism to identify which bookings are at risk until the guest simply does not show up. The challenge is not a lack of data; hotels capture rich booking signals at reservation time. The challenge is converting those signals into timely, accurate cancellation predictions that enable proactive intervention before revenue is lost.

Proposed Solution

A machine learning-powered Cancellation Prediction System that scores every new booking at the time of reservation with a cancellation risk probability (0–100%), identifies the top drivers behind each at-risk booking, and feeds these scores into a real-time dashboard for revenue managers to trigger targeted retention actions — non-refundable upsells, proactive outreach, or tightened deposit requirements — before the cancellation occurs.

Business Impact

- **Revenue protection:** Reduce cancellation-driven revenue loss from 15% of net revenue by enabling early intervention on the highest-risk bookings before arrival
- **Policy precision:** Replace blanket cancellation policies with data-driven, segment-specific policies calibrated to actual risk (corporate vs. online vs. offline channels)
- **Occupancy optimisation:** Enable more confident overbooking strategies based on predicted cancellation volumes, improving room utilisation during peak and shoulder seasons

Key Milestones

Milestone	Target
Model development & validation (RapidMiner)	Completed ✅
Production model integration (PMS / booking platform)	Q3 2026
Revenue manager dashboard live	Q4 2026

Milestone	Target
Policy automation layer	Q1 2027
Model retraining pipeline	Q2 2027

Success Metrics

Tier	Metric	Baseline	Target
North Star	Cancellation-driven revenue recovered per quarter (USD)	Not tracked	Measurable and growing
Primary	At-risk booking prediction accuracy (test set)	Random: 33% base rate	≥ 87% (Random Forest benchmark)
Primary	False positive rate (non-cancellers flagged as at-risk)	—	≤ 15%
Secondary	Cancellation rate after intervention	33% (industry)	≤ 25% within 12 months
Secondary	Revenue manager time saved per week on manual analysis	Days	Hours
Health	Model score latency per booking	—	< 2 seconds
Health	Model accuracy on monthly retraining cycle	—	≥ 85% test accuracy

2. Problem Definition

2.1 Customer Problem

- **Who:** Revenue managers, front-of-house operations leads, and booking platform product managers at mid-scale to premium hotels with 100–2,000 rooms — operating with a mix of online (64%), offline (29%), and corporate (< 1%) booking channels
- **What:** They cannot identify which of their incoming bookings are likely to cancel early enough to take meaningful action. By the time a cancellation occurs, the revenue window is closed — the room may go unsold, overbooking buffers are miscalibrated, and operational planning (staffing, F&B, housekeeping) is disrupted
- **When:** Continuously — at reservation creation (lead time up to 475 days in advance), during the booking window, and in the 7-day pre-arrival window where last-minute cancellations spike
- **Where:** Hotel Property Management Systems (PMS), online travel agency (OTA) dashboards, and internal revenue management tools

- **Why:** The cancellation decision is driven by 16+ interacting variables — lead time, price, arrival timing, stay composition, prior behaviour — with non-linear relationships that human analysts cannot manually synthesise at scale across thousands of bookings per month
- **Impact:** At a 33% industry cancellation rate, a 200-room hotel averaging $150/night and 70% occupancy loses approximately $7,700 per night in unrecovered revenue from cancellations alone. Even a 5 percentage point reduction in cancellation rate (33% → 28%) would recover $1,167 per night, or ~$425,000 annually.

2.2 Why Rule-Based Systems Fail — The Case for ML

Simple business rules (e.g., “flag bookings with lead time > 90 days as at-risk”) cannot solve this problem for three structural reasons:

1. **Non-linear interactions between features:** A booking with a 150-day lead time combined with zero special requests, a high room price, and the customer’s first booking is very different from one with the same lead time but multiple special requests, a loyalty profile, and a mid-range room rate. Rules cannot model these combinations. The Random Forest model confirmed this: `lead_time` alone has a weight of 0.22, but `avg_price_per_room` (0.20), `arrival_date` (0.15), and `no_of_week_nights` (0.13) each contribute meaningfully — and their interaction, not their individual values, drives prediction accuracy.

2. **Seasonal and temporal complexity:** Bookings are not evenly distributed — 40% occur in months 8–10 with a 500% volume spike over the base month. Cancellation rates also vary seasonally. A static threshold rule set during high season will misfire badly in low season. ML models can learn these temporal patterns from historical data automatically.

3. **Rare event detection (repeat vs. new guests):** 99.7% of guests book only once, making repeat-guest cancellation behaviour statistically marginal in any rule-based system. Yet prior cancellation history (`no_of_previous_cancellations`) is a significant predictive signal the model can weight appropriately even at low frequency.

Why ensemble methods specifically: Random Forest achieved the best generalisation in this project (87.94% test accuracy vs. 83.20% for a single Decision Tree) because bagging across 100+ trees reduces the overfitting that plagued the single Decision Tree (99.65% train / 83.20% test — a 16.5 percentage point gap). For a production deployment where we score new, unseen bookings, generalisation is paramount.

2.3 Input Data Signals — Why ML Pattern Recognition Is Required

The prediction model operates on structured booking event data. While the data is tabular (not raw text or images), the challenge is that cancellation behaviour emerges from **complex multi-dimensional patterns** that defy manual heuristics:

Feature	Type	Predictive Role
`lead_time`	Numeric	#1 predictor (weight 0.22–0.30) — longer advance

Feature	Type	Predictive Role
		bookings cancel more
`avg_price_per_room`	Numeric	#2 predictor (weight 0.15–0.20) — higher prices correlate with higher cancellation
`arrival_date` / `arrival_month`	Temporal	Seasonal cancellation patterns; peak season vs. shoulder
`no_of_week_nights` / `no_of_weekend_nights`	Numeric	Stay composition affects commitment level
`no_of_special_requests`	Numeric	More requests = more engaged guest = lower cancellation risk
`type_of_meal_plan`	Categorical	Meal commitment signals intent to stay
`room_type_reserved`	Categorical (encoded)	Room type preference correlates with segment behaviour
`market_segment_type`	Categorical	Online (64% of bookings) vs. offline vs. corporate risk profiles differ
`repeated_guest`	Binary	Repeat guests (0.3% of base) have very different cancellation behaviour
`no_of_previous_cancellations`	Numeric	Prior cancellation history is a strong behavioural predictor
`required_car_parking_space`	Binary	Parking need signals local/committed guest

A rule-based analyst examining 9,069 bookings would need to manually define threshold rules for each of these 16 variables and their interactions — an exponentially complex task with no principled way to weight their relative importance. ML ensemble methods (Random Forest) learn these weights from data, producing a calibrated probability score per booking.

Current data limitation and unstructured signal roadmap:

The V1.0 model uses only structured tabular inputs. However, hotel operations generate substantial unstructured data that carries high-value cancellation intent signals — signals that pattern-matching and keyword rules cannot reliably extract:

Unstructured Source	Format	Why Rules Fail	ML Opportunity
OTA / direct channel message history	Free-text (email, chat)	“Is it possible to change my dates?” is	LLM-based intent classification on pre-

Unstructured Source	Format	Why Rules Fail	ML Opportunity
		not a cancellation keyword but combined with high lead time and price, it is a strong intent signal	stay guest communications
PMS booking modification notes	Free-text fields	Notes like “guest called to query room” differ in risk from “guest requested early check-in” — no rule distinguishes these reliably	NLP sentiment + intent tagging on modification events
Post-stay OTA review text	Free-text (Booking.com, TripAdvisor)	Verbatim review sentiment correlates with repeat-guest cancellation behaviour — negative past reviewers cancel future bookings at higher rates	Sentiment scoring fed into the repeat-guest feature
Call centre transcripts	Voice → text (ASR output)	Spoken intent markers (hesitation, “thinking about cancelling”) require contextual language understanding that keyword spotting misses	Voice-to-text + LLM classification on cancellation-risk calls

Incorporating these unstructured signals in V2.0 would create a **multimodal prediction architecture** — tabular model for baseline scoring + LLM-based intent signals for high-uncertainty bookings — significantly improving precision on the hardest-to-predict cases (mid-range lead time, mid-range price bookings where the tabular model is least confident).

2.4 Differentiation from ChatGPT, Copilots, and General AI

A natural question: *“Can’t a revenue manager just export booking data to ChatGPT and ask which bookings will cancel?”* This approach fails in a hotel operations context for four reasons:

1. **No model scoring of new bookings in real time:** ChatGPT and Copilot are conversational tools — they cannot be embedded in a booking pipeline to automatically score each reservation at creation time without human initiation. A production

cancellation prediction system must score bookings at the millisecond a reservation is confirmed in the PMS.

2. **No probabilistic calibration:** General AI tools produce qualitative assessments, not calibrated probability scores. Revenue managers need a quantified risk score (e.g., “73% cancellation probability”) to triage interventions — not a paragraph that says “this booking seems risky.” Policy decisions (deposit requirements, overbooking buffers) require numeric outputs.

3. **No training on proprietary historical data:** ChatGPT has no access to a hotel’s own booking history, guest profiles, or seasonal patterns. A Random Forest trained on 9,069 hotel-specific records learns the INN Hotels Group’s own cancellation patterns, market mix, and room-type behaviour — not generic hospitality knowledge.

4. **Cannot retrain on new performance data:** As booking patterns shift (post-pandemic travel recovery, OTA algorithm changes, seasonal drift), the prediction model must retrain on new data. ChatGPT cannot be fine-tuned on operational booking data in a repeatable, scheduled pipeline.

2.5 Defensible Competitive Moat

The Cancellation Prediction System’s long-term value compounds through three mechanisms:

1. **Proprietary data flywheel:** Each booking scored by the model, combined with its actual outcome (cancelled/not), becomes a new training record. The model improves continuously as it accumulates hotel-specific cancellation ground truth — data that external tools (ChatGPT, generic PMS add-ons) can never access.

2. **Policy feedback loop:** As interventions (targeted deposits, loyalty incentives) are applied to high-risk bookings and outcomes are tracked, the system learns which interventions actually reduce cancellation for which guest profiles — creating a second-order optimisation layer that no manual system can replicate.

3. **Operational integration depth:** Once embedded in the PMS workflow (automated alerts, revenue manager dashboard, booking modification triggers), switching costs are high. A new system would need to replicate not just the model but the integrated workflow, historical training data, and calibrated risk thresholds.

2.6 Market Context

- **Hotel industry cancellation rate:** 33% (industry standard, validated in dataset: 2,971 cancelled out of 9,069 total = 32.7%)
- **Revenue loss from cancellations:** Up to 15% of net revenue per property
- **Digital booking dominance:** 64% of bookings through online channels — increasing the urgency of algorithmic tools as human channel management becomes insufficient
- **Growing market:** Online hotel booking market projected at $174B by 2026 (Statista); predictive analytics in hospitality growing at 8.2% CAGR

- **Why now:** The shift to digital bookings, combined with OTA cancellation-friendly policies, has structurally increased cancellation rates versus the pre-digital era. Hotels without predictive tools are competing with a growing information disadvantage

---

3. Solution Overview

3.1 What We're Building

A deployed machine learning system — built on a Random Forest classifier trained on INN Hotels Group booking data — that integrates with the hotel's Property Management System to automatically score every new booking with a cancellation risk probability at the point of reservation. The system surfaces risk scores through a revenue manager dashboard, highlights the top 3 drivers behind each at-risk booking, and enables configurable intervention rules (e.g., auto-apply non-refundable rate for bookings with > 70% cancellation probability and lead time > 60 days).

3.2 User Flows

Flow 1 — Real-Time Booking Scoring

> Guest completes reservation on OTA / direct channel → Booking record created in PMS (16 feature values captured) → PMS sends booking event to Prediction API → Random Forest model scores booking (cancellation probability 0–100%) → Score + top 3 feature drivers written back to PMS booking record → Revenue manager dashboard updated in < 2 seconds → If score > configurable threshold (default 70%): alert triggered to revenue manager queue

Flow 2 — Revenue Manager Intervention

> Revenue manager opens daily at-risk queue (bookings sorted by risk score) → For each high-risk booking: sees risk score, top drivers (e.g., "Lead time 145 days, Price $280/night, No special requests"), and recommended action → Selects intervention: offer non-refundable upgrade, send personalised pre-stay communication, request deposit confirmation, flag for overbooking buffer → Intervention logged and linked to booking outcome for model feedback loop

Flow 3 — Policy Automation (P1)

> Revenue manager sets automation rules (e.g., "For bookings with score > 80% AND lead time > 90 days: automatically switch to non-refundable rate and send email") → Rule engine applies policy to qualifying new bookings without manual review → All automated actions logged with booking ID for audit and outcome tracking

Flow 4 — Model Retraining and Monitoring

> Monthly: completed bookings (with final cancellation/not outcome) exported from PMS → Model retraining pipeline runs on updated dataset (original + new) → New model evaluated against holdout set (target: ≥ 85% test accuracy) → If accuracy ≥

threshold: new model promoted to production → If accuracy drops below threshold: alert sent to data team; prior model remains active

Flow 5 — Reporting and Policy Review

Weekly: Revenue manager views cancellation prediction accuracy report (predicted vs. actual) → Monthly: Model performance dashboard shows accuracy trend, precision, recall, and revenue impact of interventions → Quarterly: Policy review — adjust cancellation policy stringency by market segment based on risk profile data

3.3 In Scope

Feature	Priority	Description
Real-time booking risk scoring API	P0	Scores every booking at creation; returns probability 0–100% + top 3 drivers
Revenue manager risk dashboard	P0	Ranked list of at-risk bookings, risk score, drivers, intervention history
At-risk booking alerts	P0	Configurable threshold alerts pushed to revenue manager queue
Model performance monitoring	P0	Daily accuracy tracking; alert if performance drops below 85%
Feature importance explanations	P0	Per-booking driver display (“Lead time is the primary risk factor for this booking”)
Policy automation rules engine	P1	Configure auto-actions for bookings above a risk threshold
Pre-stay intervention campaign hooks	P1	Integration with CRM/email platform to trigger targeted communications
Overbooking recommendation engine	P1	Recommended overbooking buffer per arrival date based on predicted cancellation volume
Monthly model retraining pipeline	P1	Automated retrain on accumulated outcomes; accuracy gate before promotion
Cancellation driver analytics	P2	Segment-level report: which features drive cancellations most by market segment /

Feature	Priority	Description
		season
A/B testing for interventions	P2	Test whether specific interventions (non-refundable upgrade offer) reduce cancellation for high-risk segments

3.4 Out of Scope

- Natural language booking ingestion (voice reservations, chatbot bookings) — Phase 2
- Guest sentiment analysis from reviews / post-stay surveys — Phase 2
- Dynamic pricing integration (using cancellation risk to adjust pricing in real time) — Phase 3
- Multi-property aggregated models — Phase 3

3.5 MVP Definition

- **Core Features:** Real-time scoring API, risk dashboard, configurable alerts, feature importance display
- **Success Criteria:** Model achieves ≥ 85% test accuracy on live data; dashboard loads at-risk queue in < 3 seconds; scoring latency < 2 seconds per booking
- **MVP Target:** Q4 2026
- **Learning Goals:** Validate that revenue managers act on risk scores (dashboard engagement > 3x/week); validate that intervention rate on flagged bookings is > 50%; measure whether intervened bookings have lower actual cancellation rate

---

4. User Stories & Requirements

4.1 User Stories

Story 1 — Early Risk Identification As a **hotel revenue manager**, I want to **see which bookings have a high cancellation probability at the time of reservation**, So that **I can take proactive action — like requesting a deposit or offering a non-refundable upgrade — before the revenue window closes**.

Acceptance Criteria: - [ ] Every booking is scored within 2 seconds of creation in the PMS - [ ] Risk score (0–100%) and top 3 contributing feature drivers displayed per booking - [ ] Dashboard shows all bookings above configurable risk threshold (default: 70%) - [ ] Score history available for each booking (score at creation, score at 30 days, score at 7 days pre-arrival)

Story 2 — Policy Calibration by Segment As a **hotel GM setting cancellation policy**, I want to **understand which market segments and booking patterns drive the most cancellations**, So that **I can apply tighter deposit requirements to high-risk segments without penalising reliable bookers**.

Acceptance Criteria: - [ ] Segment-level cancellation driver report available (online vs. offline vs. corporate vs. aviation) - [ ] Cancellation rate by lead-time band and price range shown in weekly report - [ ] Model feature importance scores accessible to non-technical users with plain-language labels

---

Story 3 — Overbooking Optimisation As an **operations manager planning room inventory**, I want to **know the predicted cancellation volume for each upcoming arrival date**, So that **I can set a justified overbooking buffer — not an arbitrary number — and maximise occupancy without overselling**.

Acceptance Criteria: - [ ] Per-arrival-date cancellation forecast generated 30, 14, and 7 days in advance - [ ] Forecast includes confidence interval (e.g., "12–18 cancellations expected for October 15th arrival") - [ ] Actual vs. predicted cancellation volume tracked per arrival date for forecast calibration

---

Story 4 — Intervention Effectiveness Measurement As a **revenue manager wanting to know if my actions are working**, I want to **track whether bookings I intervened on had lower actual cancellation rates than similar bookings I didn't intervene on**, So that **I can refine my intervention strategy and justify the cost of incentives offered to at-risk guests**.

Acceptance Criteria: - [ ] Every intervention logged with booking ID, intervention type, date, and eventual booking outcome - [ ] Weekly report shows intervention group vs. control group cancellation rate comparison - [ ] System flags which intervention types have the highest ROI (cancellations saved per $X cost)

4.2 Functional Requirements

ID	Requirement	Priority	Notes
FR1	Model must score all new bookings in real time (< 2 seconds) via REST API	P0	Core production requirement
FR2	Score must include cancellation probability (%) + top 3 feature drivers with plain-language labels	P0	Explainability for revenue manager trust
FR3	Dashboard must display bookings ranked by risk score	P0	Actionable prioritisation

ID	Requirement	Priority	Notes
	with configurable threshold filter		
FR4	System must alert revenue manager when new bookings exceed risk threshold	P0	Proactive notification
FR5	Model accuracy must be monitored daily; alert fired if test accuracy drops below 85%	P0	Production reliability
FR6	Model must be retrained monthly on updated booking outcomes	P1	Continuous improvement
FR7	Policy automation engine must support rule creation: IF score > X AND feature Y > Z THEN action A	P1	Scaled intervention without manual review
FR8	System must log all interventions with booking ID and outcome for effectiveness tracking	P1	ROI measurement
FR9	Per-arrival-date cancellation volume forecast must be generated at 30/14/7-day horizons	P1	Overbooking optimisation
FR10	All model decisions must be auditable: feature values used at scoring time logged per booking	P0	Compliance and debugging
FR11	Dashboard must be accessible to non-technical users with no ML background	P0	Revenue manager usability requirement
FR12	System must handle new bookings from all market segments: Online, Offline,	P0	Full booking channel coverage

ID	Requirement	Priority	Notes
	Corporate, Aviation, Complementary		

4.3 AI System Capabilities and Autonomy Boundaries

Operates autonomously (no human approval): - Scoring each new booking with cancellation probability within 2 seconds of creation - Populating risk score and top feature drivers in the dashboard - Generating arrival-date cancellation forecasts on the 30/14/7-day schedule - Triggering alerts when bookings exceed threshold - Applying pre-approved policy automation rules (e.g., auto switch to non-refundable rate) - Monthly retraining pipeline execution and accuracy evaluation

Requires human approval: - Promoting a retrained model to production (human must review accuracy report and approve) - Changing the risk score threshold that triggers automated actions (revenue manager decision) - Any automated guest communication (deposit request, rate change, upgrade offer) must be reviewed or pre-approved by revenue manager for first 90 days - Overbooking buffer recommendations are advisory only — operations manager must confirm before applying to inventory

Error handling and safeguards: - If PMS sends incomplete booking data (< 12 of 16 features populated): booking is scored with available features and flagged as “partial data — lower confidence”; revenue manager notified - If model API is unavailable: bookings are queued and scored in batch when service restores; no booking blocks on scoring failure - If model accuracy drops below 80% on rolling 7-day actuals: automated actions paused; human-only review mode activated; data team alerted

4.4 Non-Functional Requirements

- **Performance:** Scoring API response < 2 seconds P99; dashboard query for at-risk queue < 3 seconds; no booking creation blocked by scoring latency
- **Scalability:** Handle 5,000 bookings/day at MVP; 50,000/day for multi-property expansion
- **Reliability:** 99.9% uptime for scoring API; asynchronous scoring queue ensures no booking is lost if API has transient failure
- **Security:** Booking data encrypted in transit (TLS 1.3) and at rest; no guest PII exposed in model features or logs; GDPR-compliant data retention policy
- **Explainability:** Every risk score accompanied by top 3 feature contributors in plain English (e.g., “High lead time,” “No special requests,” “High average price”) — black-box outputs unacceptable for revenue manager trust
- **Compliance:** GDPR Article 22 compliance — automated decisions that affect guest (e.g., rate change) must have human review capability and guest right-to-explanation pathway

---

5. Model Selection and Performance

5.1 Model Comparison

Four models were trained and evaluated on INN Hotels Group booking data (9,069 records; 70/30 train/test split):

Model	Train Accuracy	Test Accuracy	Test Recall	Test Precision	Overfitting Gap
Decision Tree	99.65%	83.20%	76.77%	73.23%	16.5 pts — HIGH
Decision Tree (Pruned)	95.64%	84.34%	76.09%	76.09%	11.3 pts
Random Forest	**98.47%**	**87.94%**	**77.10%**	**84.71%**	**10.5 pts — BEST**
Random Forest (Pruned)	87.32%	85.04%	66.22%	84.77%	2.3 pts

5.2 Recommended Production Model: Random Forest (Unpruned)

Rationale: Random Forest achieves the highest test accuracy (87.94%) and the best test precision (84.71%) — minimising false positives (non-cancellers incorrectly flagged and subjected to unwanted deposit requests or rate changes). For hotel operations, false positives carry a direct guest satisfaction cost; high precision is therefore the primary model selection criterion alongside accuracy.

The pruned Random Forest reduces overfitting (train/test gap narrowed to 2.3 points) but sacrifices 11.9 points of recall (66.22% vs. 77.10%) — meaning significantly more actual cancellations go undetected. This trade-off is unfavourable for a revenue-protection use case.

5.3 Key Feature Importance (Random Forest)

Rank	Feature	Weight	Business Interpretation
1	lead_time	0.22	Bookings far in advance cancel more — longer commitment horizon = higher risk
2	avg_price_per_room	0.20	Higher-priced rooms cancelled more — price sensitivity triggers last-minute reconsideration
3	arrival_date	0.15	Day-of-month patterns; month-end

Rank	Feature	Weight	Business Interpretation
			and specific holiday periods have distinct cancellation profiles
4	no_of_week_nights	0.13	Longer mid-week stays cancel more — possibly corporate itinerary changes
5	arrival_month	0.11	Seasonal pattern; months 8–10 peak volume but stable cancellation ratio; shoulder months vary
6	no_of_weekend_nights	0.09	Weekend stays have different risk profile vs. weekday
7	type_of_meal_plan	0.08	Meal plan commitment signals intent to stay; no meal plan = less committed
8	no_of_special_requests	0.05	More special requests = more engaged guest = lower cancellation risk

6. Go-to-Market Strategy

Deployment Plan

- **Pilot (Q3 2026):** Single property deployment; model runs in "shadow mode" (scores bookings but does not trigger automated actions); revenue manager uses dashboard manually for 60 days; outcome tracking begins
- **MVP Live (Q4 2026):** Revenue manager alert system live; manual intervention logging activated; first accuracy report generated after 30-day production window
- **Policy Automation (Q1 2027):** Rules engine enabled for pre-approved automation actions (deposit request, rate switch) on highest-confidence predictions; guest communication templates approved
- **Multi-property (H2 2027):** Expand to additional properties; train property-specific models vs. shared model evaluation

Stakeholder Rollout

- **Revenue Manager:** Primary user — trained on dashboard interpretation, intervention logging, risk threshold adjustment
- **General Manager:** Weekly reporting recipient — cancellation forecast, intervention ROI, model accuracy summary
- **Front Desk / Operations:** Receives overbooking buffer recommendations (advisory); notified of high-risk arrivals for check-in preparation

---

7. Risks & Mitigations

Risk	Probability	Impact	Mitigation
Model degrades over time as booking patterns change	High	High	Monthly automated retraining pipeline; human approval gate before production promotion; accuracy monitoring dashboard
False positives erode guest satisfaction (wrongly flagged guests receive deposit requests)	Medium	High	Tune threshold upward (85%+ for automated actions); human review required for first 90 days; guest communication framed as “preference confirmation” not “deposit demand”
Revenue managers don’t trust AI scores and ignore the dashboard	Medium	Medium	Explainability layer (top 3 drivers per booking); accuracy track record displayed on dashboard; shadow-mode pilot builds trust before automation
GDPR / data privacy challenge on automated rate changes	Low	High	Human-in-loop required for any guest-facing automated action; audit log of all model decisions; guest right to explanation pathway

Risk	Probability	Impact	Mitigation
Training data bias (9,069 records may not cover all edge cases)	Medium	Medium	Confidence intervals on predictions; “partial data” flag for atypical bookings; expand training data to 3+ years of booking history in production
Model overfitting to INN Hotels Group patterns (limits generalisability to other properties)	Low	Low	Property-specific model training for each deployment; do not share model weights across properties without retraining on property-specific data

8. Metrics Framework

North Star Metric

Cancellation-driven revenue recovered per quarter Measured as: (bookings flagged as at-risk × intervention rate × actual cancellation reduction %) × avg room revenue per night × avg stay duration. Target: Quantify and demonstrate growth after each quarterly policy review cycle.

How to measure it (counterfactual method): Revenue recovered is a counterfactual metric — we cannot directly observe “what would have happened without intervention.” The correct measurement approach is a **10% holdout A/B test**: randomly route 10% of high-risk bookings to a control group (no intervention); compare actual cancellation rate between the treatment group (intervened) and control group (no intervention). The delta in cancellation rate × avg room revenue × avg stay length = revenue recovered per quarter. This must be configured at system launch and maintained continuously — without the holdout group, the North Star metric cannot be tracked.

Full Metric Hierarchy

Tier	Metric	Unit	Target
North Star	Cancellation-driven revenue recovered per quarter	USD	Growing YoY
Primary	Model test accuracy (rolling 30-day actuals)	%	≥ 85%
Primary	False positive rate	% of flagged	≤ 15% [1]

Tier	Metric	Unit	Target
	(flagged bookings that didn't cancel)		
Secondary	Actual cancellation rate post-deployment	% of bookings	≤ 25% (from 33% baseline)
Secondary	Revenue manager dashboard engagement	Sessions/week	≥ 3 sessions/week
Secondary	Intervention rate on flagged bookings	% of at-risk acted on	≥ 50%
Secondary	Intervention effectiveness	Cancellation rate (intervened vs. control)	≥ 20% relative reduction
Financial	Revenue recovered via non-refundable upsells	USD/month	Measurable and growing
Health	Model scoring latency	Seconds P99	< 2s
Health	Model retraining success rate	% of monthly runs passing accuracy gate	≥ 95%

[1] *False positive tolerance rationale: 15% false positive rate is acceptable because wrongly flagged guests receive soft interventions — upgrade offers, confirmatory pre-stay emails — not punitive actions like deposit forfeitures. The guest satisfaction cost of a false positive is low (an unrequested upgrade offer is at worst mildly surprising), while the cost of a false negative (actual cancellation missed) is a full night of lost room revenue. For automated punitive actions (mandatory deposits, non-refundable rate switches), a stricter threshold of ≤ 5% false positive rate should be applied in the policy automation engine.*

---

9. Timeline & Milestones

Milestone	Date	Deliverables	Success Criteria
Academic model development	✅	Random Forest (87.94% accuracy), feature importance report, model selection rationale	Course submission accepted
Production API development	Q3 2026	REST API wrapper for Random Forest model; PMS integration	< 2s scoring latency; 99.9% uptime in staging
Shadow-mode pilot	Q3–Q4 2026	Dashboard live;	Revenue manager

Milestone	Date	Deliverables	Success Criteria
		scores recorded; no automated actions	reviews 100% of high-risk bookings manually
MVP live	Q4 2026	Alert system; intervention logging; accuracy monitoring	30-day accuracy on live data ≥ 85%
Policy automation	Q1 2027	Rules engine; automated deposit requests; CRM hooks	Automation handles ≥ 30% of at-risk interventions
Multi-property expansion	H2 2027	Property-specific models; shared dashboard	≥ 3 properties live; accuracy maintained ≥ 85% per property

10. Team & Resources

Role	Allocation
Product / Project Owner	Sankar Kumar Palaniappan
ML Engineer (model deployment + retraining pipeline)	1 × 100%
Backend Engineer (API + PMS integration)	1 × 100%
Frontend Engineer (dashboard)	1 × 50%
Revenue Manager (primary user + pilot feedback)	0.2 FTE (5 hrs/week)
Data Engineer (booking data pipeline)	0.5 FTE

Infrastructure Budget (Monthly):

Category	Monthly Cost
Cloud hosting (model API + dashboard)	~$300
Data storage (booking records + model logs)	~$100
Retraining compute (monthly batch job)	~$50
CRM / email integration (per property)	~$200
Total	**~$650/mo**

Revenue justification: Even recovering 10 room-nights/month from interventions at $150 avg rate = $1,500/month revenue recovered — 2.3× infrastructure cost.

---

11. Open Questions

1. **PMS integration complexity:** Which PMS system is the target hotel running (Opera, Protel, Mews)? Integration architecture varies significantly by system.
2. **Guest communication ownership:** Who drafts and approves the pre-stay communication templates for at-risk guests — marketing, revenue management, or front office?
3. **Data retention for model training:** How many years of historical booking data is available beyond the 9,069 records in the academic dataset? Production models benefit significantly from 3–5 years of data.
4. **Multi-property model strategy:** Should each property have its own model (better accuracy for unique patterns) or a shared model with property-specific fine-tuning (faster deployment, shared learning)?
5. **Overbooking policy ownership:** Who has authority to change the overbooking buffer based on model forecasts — revenue manager, GM, or yield management system?

---

12. Assumptions Made

- Production model is the Random Forest (unpruned) based on best test accuracy (87.94%) and precision (84.71%) — best trade-off for false positive minimisation
- The 9,069-record INN Hotels Group dataset is representative of the target property's booking patterns; production accuracy may vary with different property mix
- Revenue managers will need a 60-day shadow-mode onboarding period to build trust in AI scores before acting on them independently
- "Intervention" is defined as any revenue-manager-initiated action taken within 14 days of a high-risk booking being flagged
- GDPR compliance requires that no fully automated adverse action (rate change, cancellation fee) can be applied to a booking without human review in the loop — automated soft actions (confirmatory email, upgrade offer) are permitted with appropriate disclosure
- The $150/night average room rate and 70% occupancy assumptions used in the business case are illustrative estimates; actual figures should be substituted for the target property