COVID-19 ICU Admission Prediction Project – Requirements & Evidence Matrix (Updated)

Executive Summary

This project develops a machine learning–based early-warning system to predict ICU admissions for COVID-19 patients in Brazil using the Sírio-Libanês Hospital dataset (1,925 records, 231 features, 5 temporal windows).
It combines technical rigor with clinical relevance—bridging data science, hospital operations, and executive decision-making to optimize critical-care resource allocation during pandemic conditions.

Stakeholders:

- Product Owner: QS Academy Faculty – Data Analytics Track
- Data Scientist: Nathan Weber
- Reviewers: 2025 Cohort – Applied Machine Learning Capstone
- Use Case: Early ICU prediction and resource optimization for hospital management

Requirements–Evidence Traceability Matrix

ID	Requirement Area	Deliverable / Description	Evidence / Proof of Completion	Priority
PSR-1	Dataset Load & Integrity Check	Load and preview the Sírio-Libanês dataset. Show DataFrame shape, dtypes, column list, and head/tail.	Jupyter cell output with df.shape, df.info(), df.head()	High
PSR-2	Unified Notebook Submission	One complete .ipynb or .py file including all steps, visuals, and markdown explanations.	GitHub repo link + single notebook; README with step summary	High

PSR-3	Missing/Outlier Cleaning	Drop duplicates; analyze missingness per column; remove features with >70% missing; impute remainder.	Printed missing-value summary table before/after cleaning	High
PSR-4	Feature Encoding	Encode categorical features (e.g., AGE_PERCENTIL, binary comorbidities).	Value counts and column listings before/after encoding	High
PSR-5	Standardization	Scale numeric features (Min-Max or StandardScaler).	Transformation applied within pipeline; verified on test set	High
PSR-6	EDA & Visualization	At least five labeled charts: ICU rate over time, age distribution, comorbidity heatmap, vitals trend, lab completeness.	Matplotlib/Seaborn outputs + markdown interpretation	High
PSR-7	Temporal Analysis	ICU probability progression across 0–2h → >12h windows.	Line chart and summary statistics confirming 8.3%→50.6% increase	High
PSR-8	Modeling – Baselines	Train Logistic Regression, Random Forest, Gradient Boosting as base learners.	Pipeline outputs; printed F1, ROC-AUC, confusion matrices	High

PSR-9	Hyperparameter Tuning	Nested 5-fold CV with GridSearchCV for each model (optimize F1).	Code logs and tables of best params, F1 per fold	Critical
PSR-10	Ensemble & Stacking	Combine tuned models in a stacking ensemble with logistic meta-learner.	Ensemble training results with F1 ≥ 0.9 on hold-out set	High
PSR-11	Recall Optimization	Adjust thresholds / class weights to minimize false negatives (missed ICU cases).	Threshold sweep plots; recall vs. precision curve	High
PSR-12	Feature Importance & Interpretability	Present top 10 predictive features; discuss clinical significance.	Bar chart of feature importances; clinical notes	High
PSR-13	Model Evaluation Metrics	Report accuracy, F1, ROC-AUC, precision, recall, specificity.	Classification reports and confusion matrices printed	High
PSR-14	Executive Presentation	3-page deck: Executive Summary, Clinical Deep Dive, Operational Insights.	PowerPoint/PDF with visuals (KPI cards, trend lines, Sankey, heatmap)	High
PSR-15	Power BI Dashboard (Optional)	OPTIONAL: Publish interactive dashboard mirroring executive deck	PBIX file or screenshots (if implemented). Not required for grading or	Optional

		structure for visualization or demonstration purposes.	completion.	
PSR-16	Research Report (Technical)	2000-word report including Abstract, Methods, Results, Discussion.	PDF export with tables of metrics and hyperparameters	High
PSR-17	Repository & Documentation	GitHub repo with commit history, code, report, and presentation exports.	Verified commit log, structured folders, README.md	High

Deliverable Summary

Deliverable	Description	Format
Exploratory Data Analysis	Comprehensive temporal, demographic, and lab feature analysis	Jupyter Notebook / PDF
ML Model Development	Tuned Logistic, RF, GB + Stacking Ensemble	Python (.ipynb / .py)
Executive Presentation	Narrative deck for hospital leadership	PowerPoint / PDF
Research Report	Technical report with statistical validation	PDF
Power BI Dashboard (Optional)	Interactive analytics dashboard for demonstration (optional deliverable)	PBIX (Optional)
GitHub Repository	Full reproducible code and documentation	Public Repo