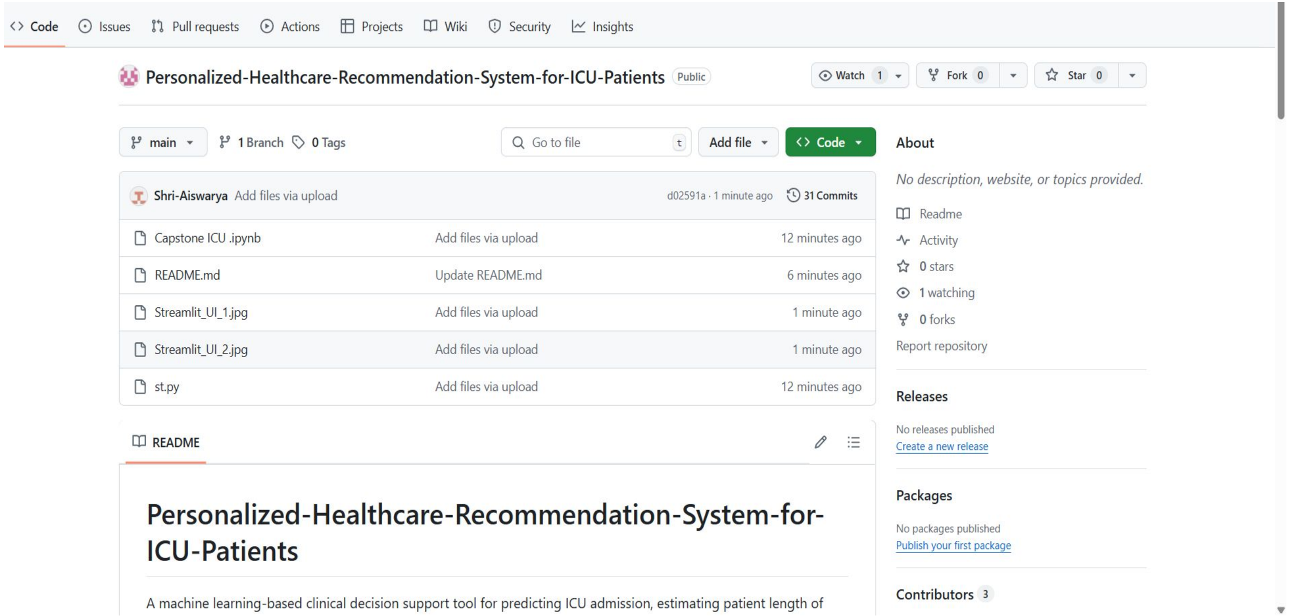
Personalized Healthcare Recommendation system for ICU Patients

TEAM D:
SHRI AISWARYA GORLE
SHASHANK RAJ GUPTA GUNTA
VENKATA NAGA SHARANYA KHANDE RAO

GITHUB REPO:

Link: https://github.com/sharanya123-khanderao/Personalized-Healthcare-Recommendation-System-for-ICU-Patients



Shri-Aiswarya Add files via upload d02591a · 1 minute ago 31 Commits

Capstone ICU .ipynb	Add files via upload	12 minutes ago
README.md	Update README.md	6 minutes ago
Streamlit_UI_1.jpg	Add files via upload	1 minute ago
Streamlit_UI_2.jpg	Add files via upload	1 minute ago
st.py	Add files via upload	12 minutes ago

Personalized-Healthcare-Recommendation-System-for-ICU-Patients

A machine learning-based clinical decision support tool for predicting ICU admission, estimating patient length of

About

No description, website, or topics provided.

Readme
Activity
0 stars
1 watching
0 forks
Report repository

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 3

WHY THIS PROJECT?

- Life-Saving Impact: Early detection of patient deterioration enables prompt, potentially life-saving interventions.
- Multimodal Data Integration: Combines structured data (vitals, labs, history) with unstructured clinical notes for a comprehensive view.
- Innovative Techniques: Leverages advanced ML & NLP to uncover hidden patterns and predict adverse events.
- Real-World Relevance: Uses the robust, publicly available MIMIC-III dataset to simulate realistic ICU scenarios.
- Scalability & Future Integration: The project pipeline is designed with scalability in mind, paving the way for real-time deployment in ICU decision-support systems.

LITERATURE SURVEY & OUR CONTRIBUTION

What others have done?

- Single-Task Focus: Most studies address either mortality prediction or length of stay—rarely both.
- Static Models: Limited use of time-series data like real-time vitals and lab trends.
- Limited Generalizability: Prior work often uses proprietary or simulated datasets.
- Minimal Deployment: Few models are integrated into clinical-facing tools.
- Lack of Multimodal Fusion: Sparse integration of structured and unstructured data.

What we have done:

- Dual Prediction Tasks: Simultaneous modeling of mortality risk and ICU length of stay.
- Predicted patient deterioration early by analyzing vital signs, lab results, and medical history.
- First-Hour Data Focus: Early vitals, demographics, and admission info for rapid predictions.
- High-Fidelity Dataset: Uses real-world, public MIMIC-III (10k) ICU data.
- Our approach uniquely integrates TF-IDF-based NLP with clinical text to predict patient mortality, demonstrating real-world applicability through a scalable and interpretable model.
- Selected Random Forest models for their balance between performance and interpretability.
- Created a Streamlit web application to make predictions accessible to end users in real-time.

DATASET OVERVIEW

Dataset: MIMIC-III 10k(Kaggle)

Link: https://www.kaggle.com/datasets/bilal1907/mimic-iii-10k
Source: MIT Lab for Computational Physiology (via PhysioNet)
Host: Kaggle
Size: 10,000 ICU patient records
Format: Structured .csv files (demographics, vitals, lab results, diagnoses, and clinical notes)
#Samples: 10,000 patients
#Raw Attributes: Over 100 columns across files, including:

- Vitals: Heart Rate, Respiratory Rate, Temperature, BP, O2 Saturation
- Demographics: Age, Gender, Ethnicity, Insurance
- Clinical: Diagnoses, Comorbidities, ICU type, LOS
- Notes: Raw clinical text from providers

Objective:

To develop predictive models for ICU patients suffering **SEPSIS** outcomes using MIMIC-III data:

WHAT IS SEPSIS?

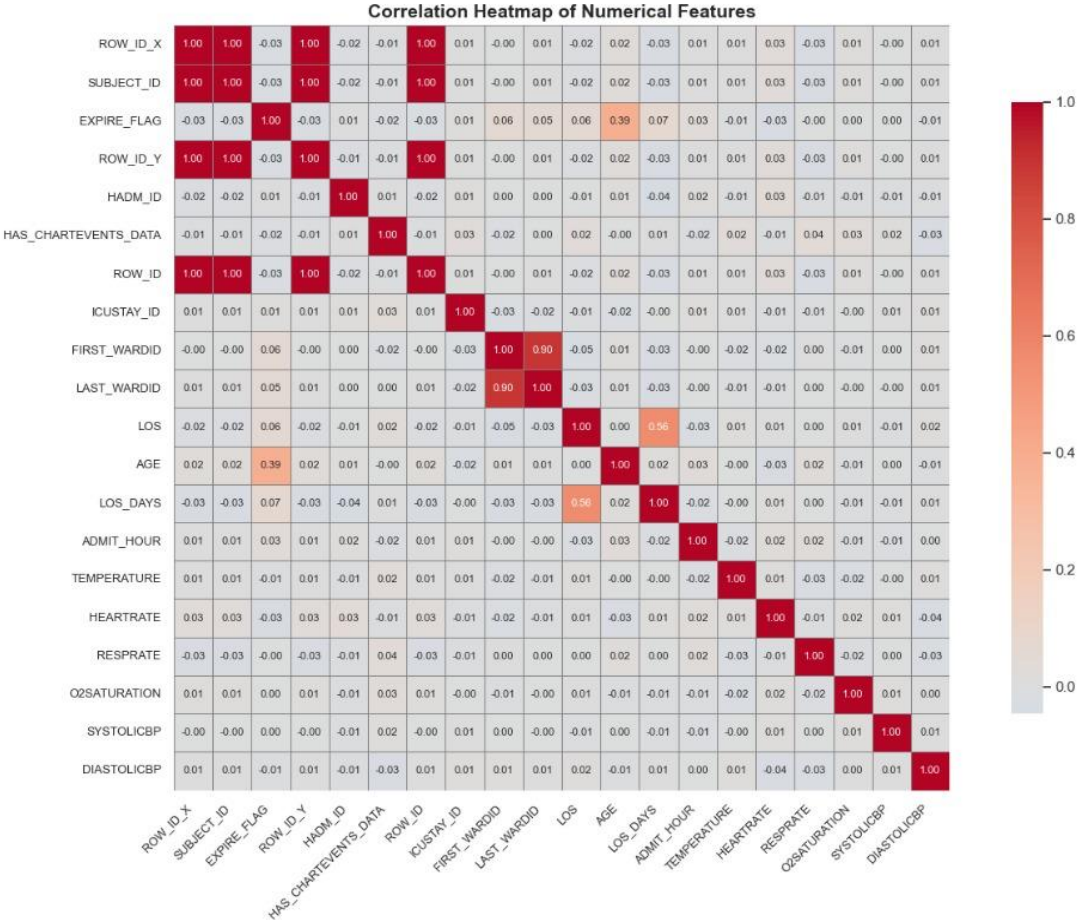
Sepsis is a life-threatening condition caused by the body's extreme response to infection, leading to organ failure. In the MIMIC-III dataset, sepsis is a critical focus as it is common in ICU patients and associated with high mortality if not treated early.

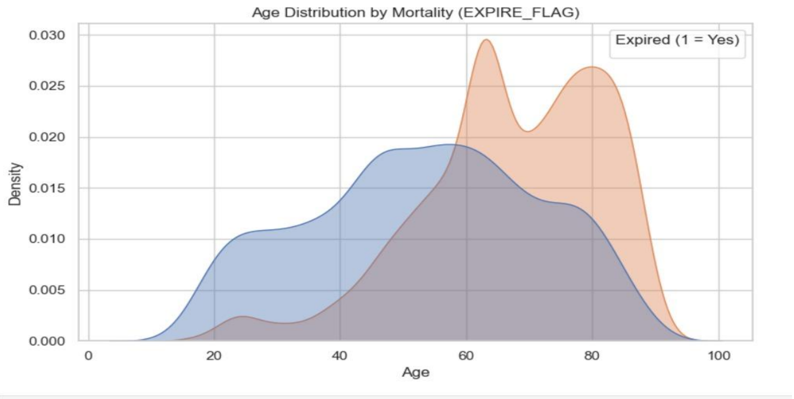
Early Deterioration Prediction (e.g., in-hospital mortality)
Risk Stratification (e.g., severity scoring, comorbidity risk)
Time-Series Modeling (e.g., trends in vitals/labs)

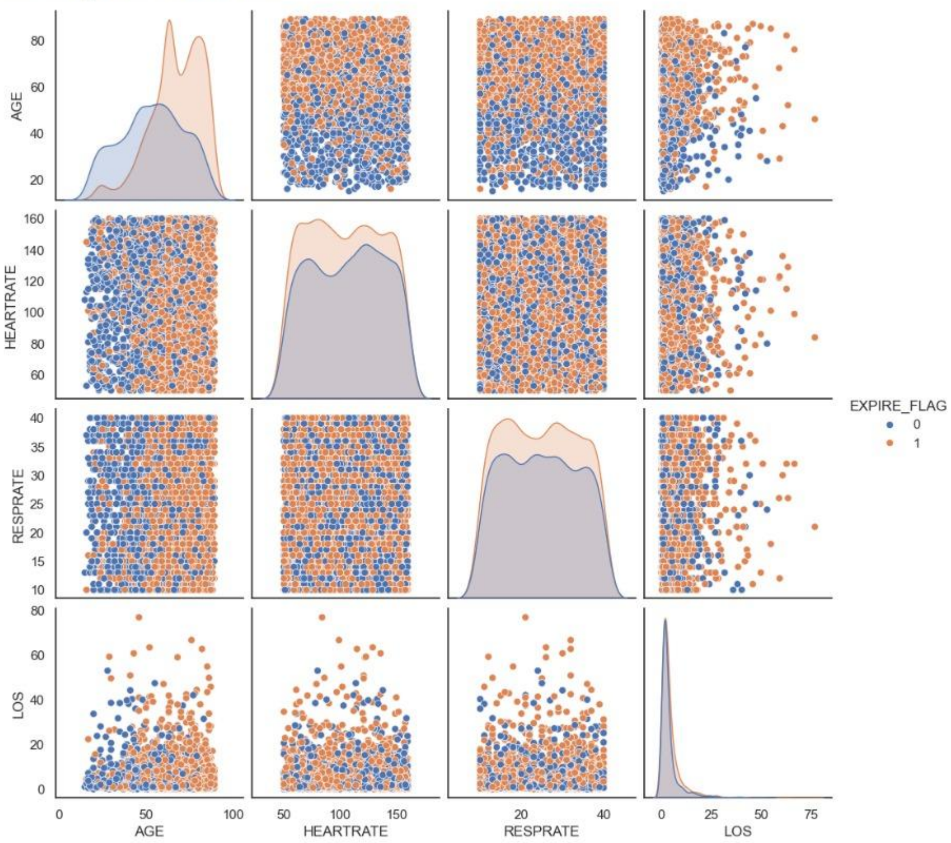
UMBC

EXPLORATORY DATA ANALYSIS

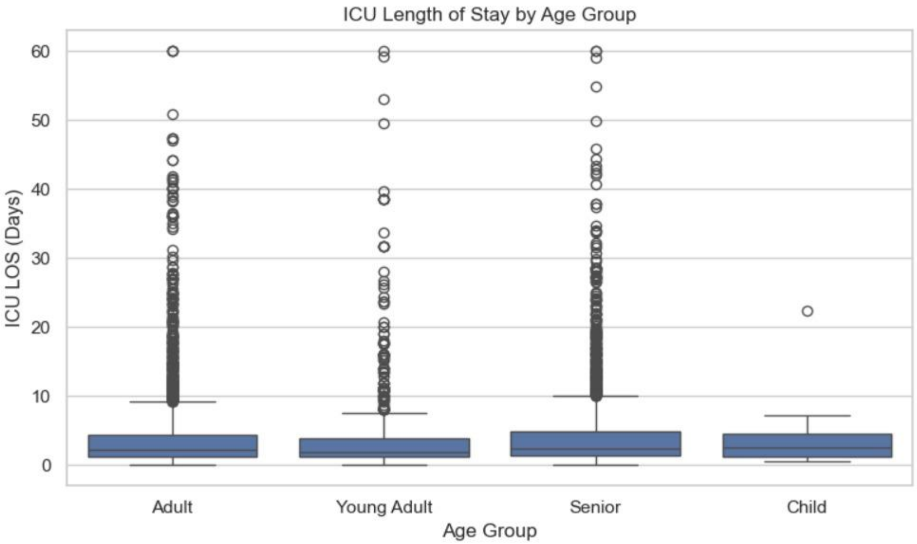
- AGE shows the strongest correlation (+0.39) with EXPIRE_FLAG, indicating that older patients are more likely to face critical outcomes—a clinically significant insight.
- Vital signs like HEARTRATE, RESPIRATORY RATE, O2SATURATION, and TEMPERATURE display low individual correlation (< 0.1) with mortality, suggesting that no single vital alone is a strong predictor.
- Despite low pairwise correlations, these vitals still hold value—when combined, they contribute to non-linear patterns captured effectively by models like Random Forest and XGBoost.
- SYSTOLICBP and DIASTOLICBP show very low correlation with the outcome, but their role in broader physiological assessment makes them essential in the feature set.
- Overall, the heatmap supports the need for multivariate analysis—outcomes are influenced by the combined effect of age, vitals, and diagnosis, rather than any isolated feature.

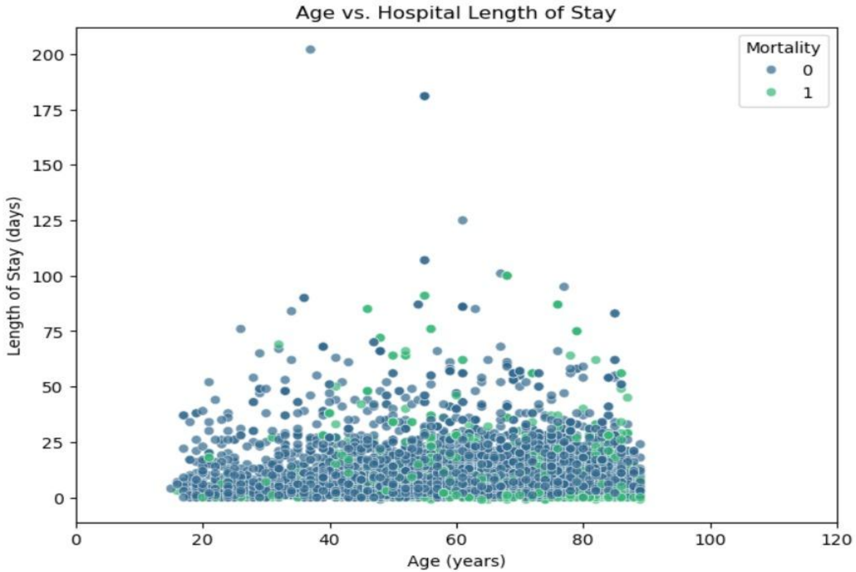

Correlation Heatmap of Numerical Features





- Age is the strongest visual predictor of mortality as, Patients aged 70 and above show a significantly higher death rate, as highlighted in the density plot, while survivors are more concentrated between 30–60 years.
- The pair plot shows overlapping distributions for vitals like HEARTRATE, RESPIRATE, and LOS, indicating that no single vital independently separates survival from death—but their combined interactions still hold predictive value.
- These visual trends confirm the need for non-linear models like XGBoost, which can learn from subtle multi-feature patterns rather than relying on linear or isolated thresholds.

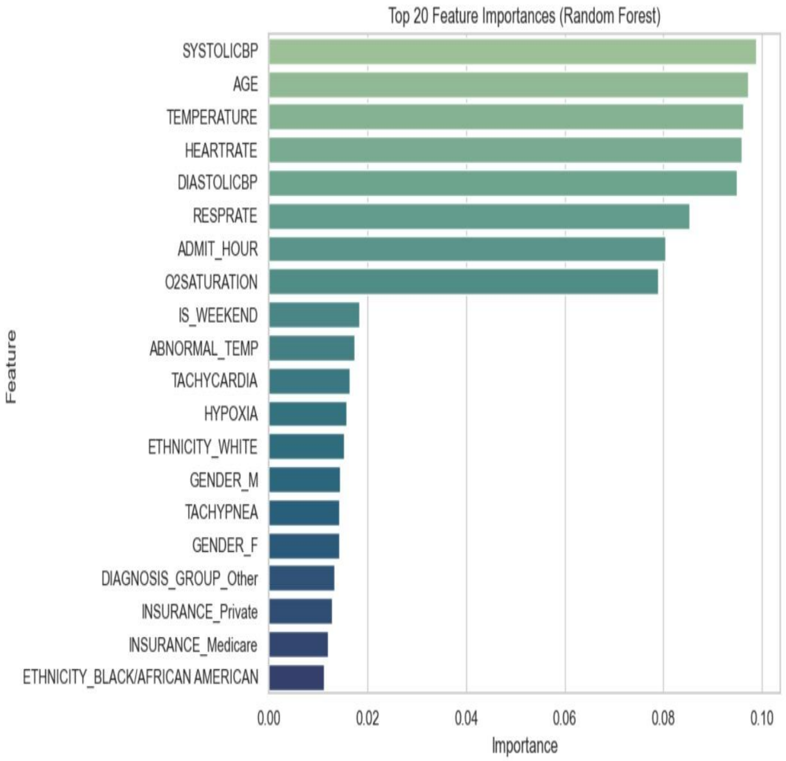




- AGE consistently shows clear patterns across visualizations, making it a highly valuable feature for identifying patient groups with distinct care needs and outcome trends.
- Estimated Length of Stay demonstrates strong practical relevance by capturing the duration of critical care, which aligns closely with patient condition severity and treatment outcomes.
- Together, these features contribute significant insight into ICU dynamics, and their inclusion in predictive models enhances the ability to forecast outcomes and optimize resource allocation with greater accuracy.

FEATURE ENGINEERING:

- Categorical variables including GENDER, INSURANCE, ETHNICITY were encoded using One-Hot Encoding for model compatibility.
- Raw features were combined with engineered features into a single pipeline for training with Random Forest.
- Feature importance plot confirms that combining raw vitals, demographics, time-based features, and engineered clinical flags created a powerful and enriched dataset for accurate model predictions.
- Top features like SYSTOLICBP, AGE, TEMPERATURE, and HEARTRATE validate both clinical relevance and data preparation strategy.



TARGET VARIABLES & FEATURE DESCRIPTION:

❑ **Target Variables**

Admission Prediction

- **Type**: Binary Classification
- **Target Variable**: ADMISSION_STATUS
- **Categories**:
  - 0 = Patient **discharged**
  - 1 = Patient **admitted**

Length of Stay Prediction

- **Type**: Regression
- **Target Variable**: LOS (Length of Stay in ICU, measured in **days**)
- **Output**: Continuous numerical value (e.g., 3.2 days, 7.0 days)

Risk Level Classification

- **Type**: Multi-Class Classification (Derived from LOS)
- **Goal**: Categorize admitted patients into clinical **risk tiers**
- **Risk Categories** (based on LOS value):
  - **Low Risk**: LOS ≤ 2 days
  - **Medium Risk**: 2 < LOS ≤ 5 days
  - **High Risk**: LOS > 5 days

Clinical notes derives using NLP techniques

- **Type:** Unsupervised Learning / Clustering
- **Target Variable:** *None* (unsupervised — no labeled output)
- **Input:** TEXT column from NOTEEVENTS.csv (unstructured clinical notes)
- **Output:** Discrete cluster labels (e.g., cluster 0, cluster 1, ..., cluster 4) assigned to each note

❑ **Key Features**

Vital Signs (from CHARTEVENTS.csv):
Heart Rate, Respiratory Rate, Temperature, Blood Pressure (SBP/DBP), SpO_2
Demographics (from PATIENTS.csv): Age, Gender, Ethnicity
Clinical Information:
Diagnoses (from DIAGNOSES_ICD.csv)
ICU type and admission source (ICUSTAYS.csv, ADMISSIONS.csv)
Comorbidities (derived from diagnosis codes)

MODEL TRAINING

❑ **Models Trained**:

- **Logistic Regression** – Used as a baseline interpretable and effective for binary classification.
- **Random Forest (Classifier & Regressor)** – Captures non-linear patterns, provides feature importance, and works well for both classification (admission/mortality) and regression (LOS).
- **XGBoost Classifier** – High-performance gradient boosting; handles class imbalance and is ideal for fine-tuned prediction.

❑ **Why Use All Three?**

Each model offers unique advantages:

- **Logistic Regression** = Simple & transparent
- **Random Forest** = Robust & explainable
- **XGBoost** = Accurate & tunable

❑ Helps **compare performance** across different algorithm types.

❑ Ensures the final model is both accurate and suitable for real-world deployment.

EVALUATION METRICS

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE	ROC AUC
LOGISTIC REGRESSION	0.67	0.66	0.78	0.71	0.706
RANDOM FOREST	0.81	0.81	0.84	0.82	0.863
XGBOOST	0.69	0.68	0.81	0.74	0.745

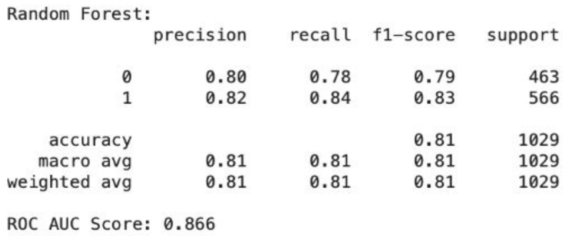
- **Accuracy**
  Measures the overall correctness of the model’s predictions
  Good general indicator, but **not sufficient alone** for imbalanced datasets.

- **Precision**
  Measures how many predicted positives were actually correct
  Important in medical settings to **avoid false positives** (e.g., wrongly admitting a patient).

- **Recall (Sensitivity)**
  Measures how many actual positives were correctly predicted
  Crucial in healthcare to catch as many high-risk (true positive) patients as possible.

- **F1-Score**
  Harmonic mean of Precision and Recall
  Best when **false positives and false negatives are both costly**, as in ICU triage.

- **ROC AUC (Receiver Operating Characteristic - Area Under Curve)**
  Indicates the model's ability to distinguish between classes
  A **threshold-independent** metric useful for binary classification.



```
Random Forest:
              precision    recall  f1-score   support

           0       0.80      0.78      0.79       463
           1       0.82      0.84      0.83       566

    accuracy                           0.81      1029
   macro avg       0.81      0.81      0.81      1029
weighted avg       0.81      0.81      0.81      1029

ROC AUC Score: 0.866
```

- **Support**
  Indicates the **number of true samples per class** (e.g., 547 admitted, 547 discharged), helping assess data balance during evaluation.
  Ensures that metrics like Precision, Recall, and F1-Score are **reliable and meaningful**, especially when classes are balanced as in your case.

- **Confusion Matrix**
  Visualizes **true positives, false positives, false negatives, and true negatives**, showing **exactly where the model succeeds or fails**.
  Helped confirm **Random Forest had fewer false negatives**, making it ideal for **high-risk ICU settings** where missing critical cases can be life-threatening.

- **Random Forest** achieved the **highest scores across all major metrics**:
  - **Accuracy**: 0.81 (highest)
  - **F1-Score**: 0.82 (highest)
  - **ROC AUC**: 0.863 (highest)
- Additionally, it maintained **strong recall (0.84)**, which is critical for capturing high-risk patients in a medical setting.
- Therefore, **Random Forest** was selected as the best model due to its **overall superior performance**, ability to **handle mixed feature types**, and **robust generalization** on unseen data.

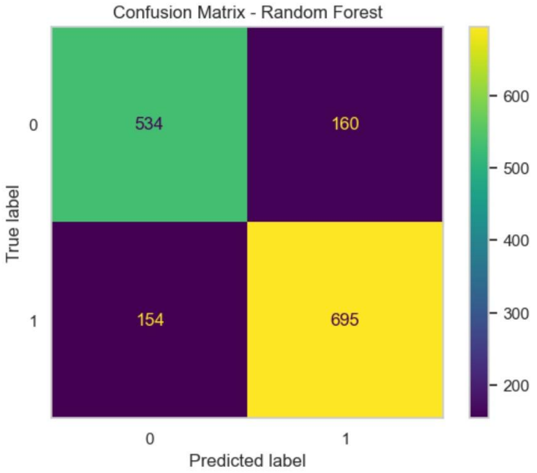
How RandomForest is used after finding it is the best one?

Predicting Admission Status

- **Goal**: Predict whether a patient will be admitted (1) or discharged (0)
- **Model Used**: RandomForestClassifier
- **Input Features**:
  - First-hour vitals: HEARTRATE, RESPRATE, O2SATURATION, etc.
  - Demographics: AGE, GENDER, ETHNICITY, etc.
  - Triage or arrival details

admission_prediction = rf_classifier.predict(patient_features)

- If admission_prediction == 1, move to step 2



Predicting Length of Stay (LOS)

- **Goal**: Estimate how many days the admitted patient will stay in the hospital
- **Model Used**: RandomForestRegressor
- **Input Features**: Same features used for admission prediction

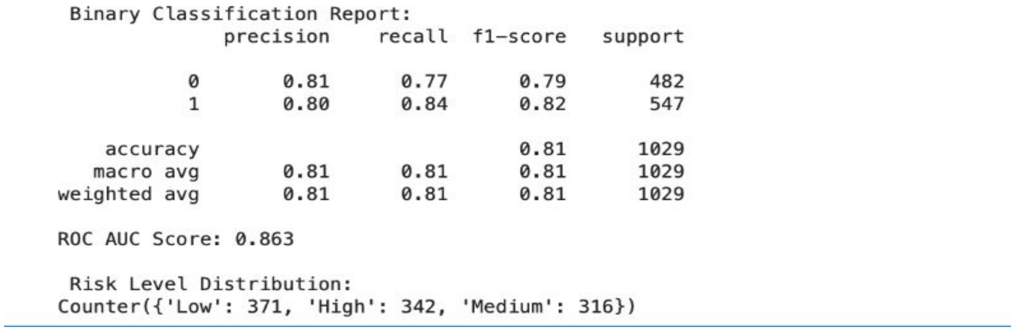
los_prediction = rf_regressor.predict(patient_features)

- Output is a **continuous number** (e.g., 3.4 → 3.4 days)

Classifying patients into Risk Levels

- We classified patients into risk levels based on their predicted probability of mortality using a **Random Forest model**.
- Then, we converted the predicted probabilities into risk levels:
  - **Low risk**: Probability < 0.4
  - **Medium risk**: 0.4 ≤ Probability < 0.7
  - **High risk**: Probability ≥ 0.7

Binary Classification Report:



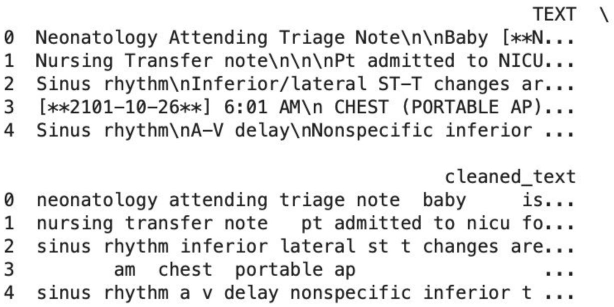
```
 Binary Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.77      0.79       482
           1       0.80      0.84      0.82       547

    accuracy                           0.81      1029
   macro avg       0.81      0.81      0.81      1029
weighted avg       0.81      0.81      0.81      1029

ROC AUC Score: 0.863

 Risk Level Distribution:
Counter({'Low': 371, 'High': 342, 'Medium': 316})
```

This classification helps us to:

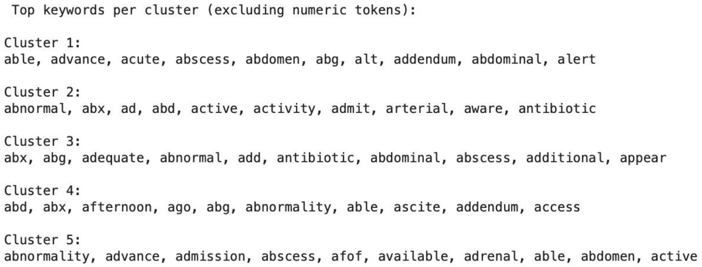
- ➢ Transforms raw probability scores into interpretable risk levels (Low, Medium, High), enabling data-driven triage and early warning for ICU patients.
- ➢ Empowers clinicians to prioritize interventions based on predicted severity rather than binary outcomes, improving care allocation and decision-making.

NLP Analysis of Clinical Notes

- We applied **Natural Language Processing (NLP)** techniques to extract hidden patterns from the **clinical notes (NOTEEVENTS.csv)** of ICU patients.
  Focus: Understand underlying themes, frequent terms, and cluster patient narratives using unsupervised learning.
- **Methodology:**
- **Text Cleaning**: Removed de-identified tags, stopwords, punctuation, and lemmatized tokens.
- **Vectorization**: Applied **TF-IDF** to convert notes into numerical vectors.
  Dimensionality Reduction: Used **Truncated SVD** to reduce feature space while retaining semantic meaning.
  Clustering: Performed **K-Means clustering (k=5)** to group similar patient notes.



```
                                              TEXT  \
0  Neonatology Attending Triage Note\n\nBaby [**N...
1  Nursing Transfer note\n\n\nPt admitted to NICU...
2  Sinus rhythm\nInferior/lateral ST-T changes ar...
3  [**2101-10-26**] 6:01 AM\n CHEST (PORTABLE AP)...
4  Sinus rhythm\nA-V delay\nNonspecific inferior ...

                                      cleaned_text
0  neonatology attending triage note  baby     is...
1  nursing transfer note   pt admitted to nicu fo...
2  sinus rhythm inferior lateral st t changes are...
3        am  chest  portable ap                  ...
4  sinus rhythm a v delay nonspecific inferior t ...
```



```
 Top keywords per cluster (excluding numeric tokens):

Cluster 1:
able, advance, acute, abscess, abdomen, abg, alt, addendum, abdominal, alert

Cluster 2:
abnormal, abx, ad, abd, active, activity, admit, arterial, aware, antibiotic

Cluster 3:
abx, abg, adequate, abnormal, add, antibiotic, abdominal, abscess, additional, appear

Cluster 4:
abd, abx, afternoon, ago, abg, abnormality, able, ascite, addendum, access

Cluster 5:
abnormality, advance, admission, abscess, afof, available, adrenal, able, abdomen, active
```

VALIDATION

Train-Test Split (80/20)

- Dataset split into:
  80% Training Set (used for model fitting & cross-validation)
  20% Test Set (held out for final evaluation only)
- Ensures unbiased evaluation of model performance on unseen data.

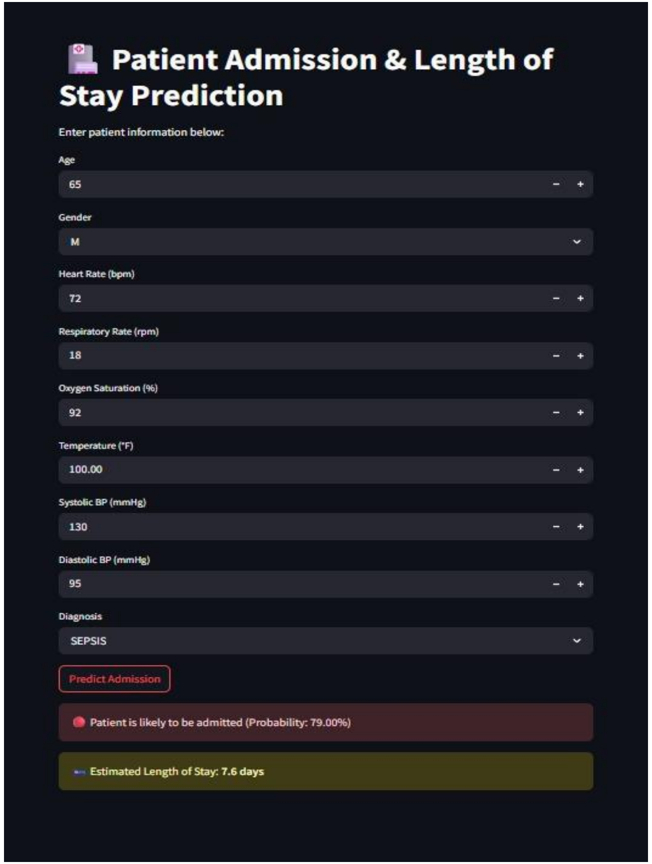
Stratified K-Fold Cross-Validation (10-Fold)

- Applied on the **training set only**
- **Stratified** → preserves class distribution (e.g., 0 = discharged, 1 = admitted)
- **10-Fold** → training done on 9 folds, validated on 1 fold, repeated 10 times
- Performance metrics (e.g., F1-Score) averaged across folds for stability

Why These Methods?

- **Stratification** ensures fair validation even with class imbalance
- **10-Fold CV** provides a **robust and reliable performance estimate**
- **Final evaluation** is done only once on the 20% hold-out test set to avoid data leakage

STREAM-LIT DEPLOYMENT:



Patient Admission & Length of Stay Prediction

Enter patient information below:

Age: 65

Gender: M

Heart Rate (bpm): 72

Respiratory Rate (rpm): 18

Oxygen Saturation (%): 92

Temperature (°F): 100.00

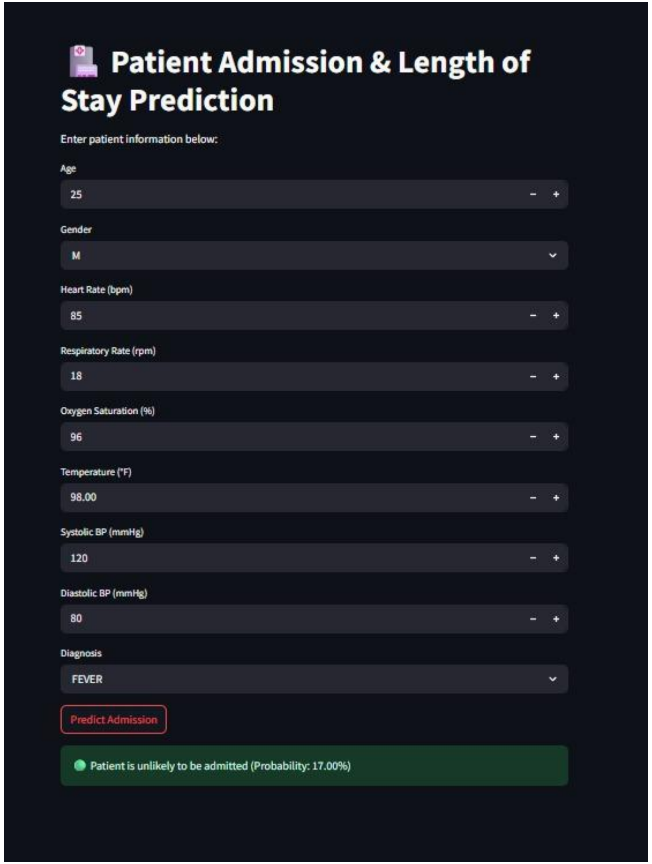
Systolic BP (mmHg): 130

Diastolic BP (mmHg): 95

Diagnosis: SEPSIS

● Patient is likely to be admitted (Probability: 79.00%)

Estimated Length of Stay: 7.6 days



Patient Admission & Length of Stay Prediction

Enter patient information below:

Age: 25

Gender: M

Heart Rate (bpm): 85

Respiratory Rate (rpm): 18

Oxygen Saturation (%): 96

Temperature (°F): 98.00

Systolic BP (mmHg): 120

Diastolic BP (mmHg): 80

Diagnosis: FEVER

● Patient is unlikely to be admitted (Probability: 17.00%)

Web Application Overview: A Streamlit-based application developed to predict patient admission status, estimated length of stay, and risk level in ICU settings using trained Random Forest models.

Interactive Features: Users can input clinical and demographic data such as age, heart rate, respiratory rate, oxygen saturation, temperature, blood pressure, and diagnosis for real-time prediction.

Real-Time Predictions: The app uses a trained Random Forest Classifier to predict admission likelihood and, if admitted, uses a Random Forest Regressor to estimate length of stay. It also classifies patients into Low, Medium, or High risk based on LOS.

User-Friendly Design: Intuitive and responsive interface with dropdowns and sliders for efficient data input and immediate feedback.

Application Impact: Demonstrates how machine learning can support real-time triage decisions, improve ICU resource planning, and enhance patient care in critical environments.

FUTURE SCOPE & REAL WORLD IMPACT

- Deploy in real-time EHR systems for live triage and ICU decision support.
- Add automated alerts based on risk level (Low, Medium, High) for proactive care.
- Enable hospital-specific model tuning to reflect local admission patterns.
- Extend to time-series modeling using LSTM or Transformer-based architectures for early warning systems.
- Implement continual learning pipelines to update models with new patient data.
- Support day-to-day hospital operations by improving resource planning and patient flow.
- Enhance public health readiness by forecasting ICU load and outbreak severity in real time.
- Assist clinicians with decision support in high-pressure or under-resourced environments.

CONCLUSION

- **Actionable Insights**: EDA and feature importance revealed critical indicators like age, systolic BP, and temperature influencing ICU admission and stay duration.
- **Strong Model Performance**: Random Forest delivered high accuracy and F1-score, outperforming Logistic Regression and XGBoost for both classification and regression tasks.
- **Two-Step Prediction Pipeline**: Successfully implemented admission prediction followed by LOS estimation and risk-level classification—mimicking real-world triage flow.
- **Streamlit Deployment**: The interactive web app allows real-time predictions from clinical inputs, supporting frontline decision-making.
- **Real-World Value**: Demonstrates how machine learning can enhance hospital operations by supporting early ICU planning, risk triage, and patient prioritization.

REFERENCES

1. Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, L.-w.H., Feng, M., Ghassemi, M.,Moody, B., Szolovits, P., Celi, L.A., & Mark, R.G. (2016). **MIMIC-III, a freely accessible critical care database.** *Scientific Data*, 3, 160035. https://physionet.org/content/mimiciii/1.4/

2. Kaggle Contributor. (2022). **MIMIC-III 10K Subset Dataset.** Retrieved from: https://www.kaggle.com/datasets/bilal1907/mimic-iii-10k

3. ZHAW School of Engineering. (2019). **Data Analysis for Mortality Prediction using MIMIC-III.** [PDF]. Available at: https://www.zhaw.ch/storage/.../DataAnalysisMortalityPrediction.pdf

4. Rajkomar, A., Dean, J., & Kohane, I. (2019). **Machine learning in medicine.** *New England Journal of Medicine*, 380(14), 1347–1358.

5. Huang, Z., Dong, W., Duan, H., & Liu, J. (2015). **A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records.** *IEEE Transactions on Biomedical Engineering*, 65(5), 956–968.

6. Huang, Y., Cai, W., & Ji, L. (2022). **Personalized Risk Prediction and Early Warning System for ICU Patients.** *Frontiers in Public Health*. https://www.frontiersin.org/articles/10.3389/fpubh.2021.818439/full

7. Lundberg, S.M., & Lee, S.-I. (2017). **A Unified Approach to Interpreting Model Predictions.** In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 4765–4774.

Thank
You