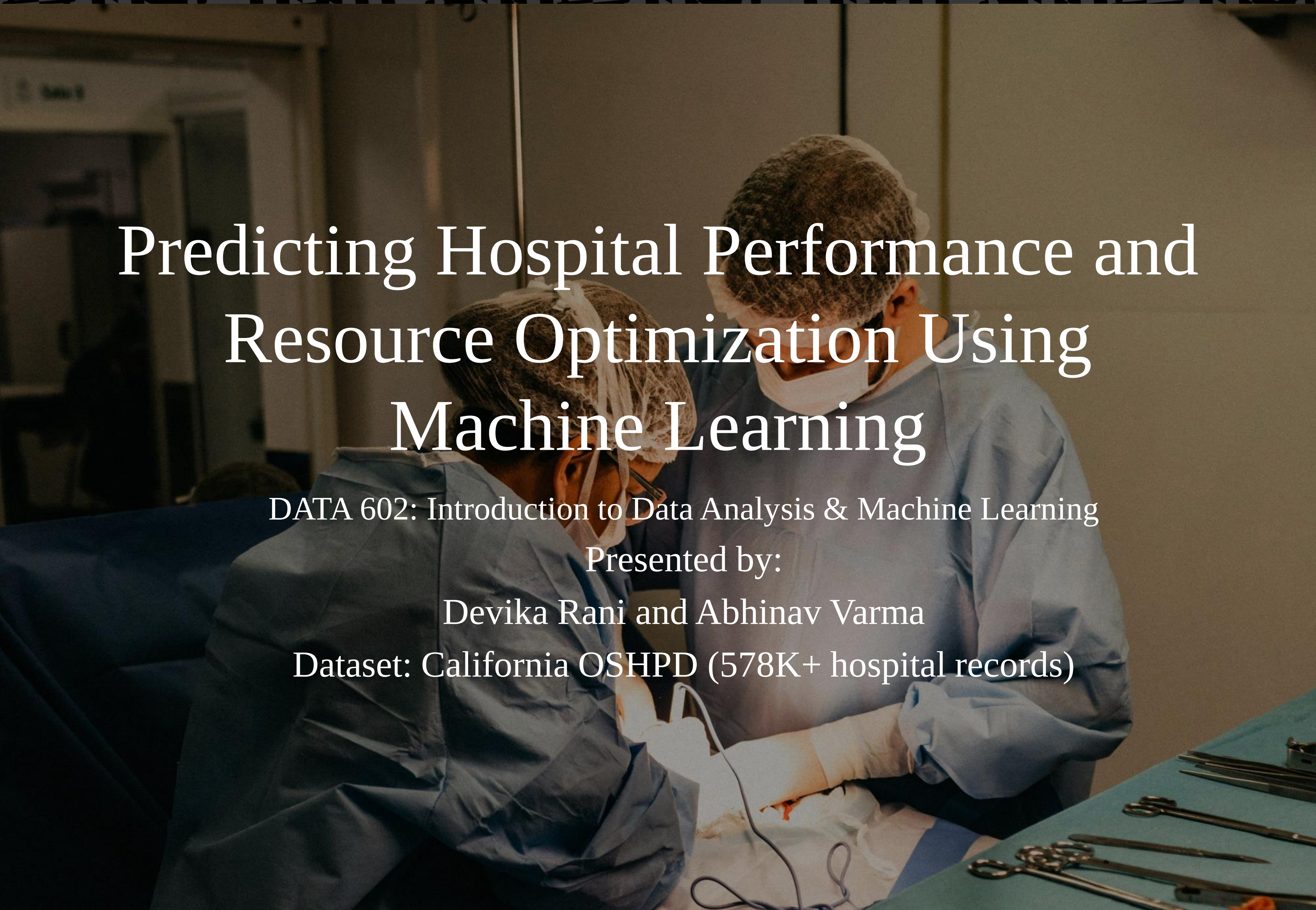
UMBC



Predicting Hospital Performance and Resource Optimization Using Machine Learning

DATA 602: Introduction to Data Analysis & Machine Learning

Presented by:

Devika Rani and Abhinav Varma

Dataset: California OSHPD (578K+ hospital records)

Project Objective

Hospitals vary in how they manage staff, beds, and patient services.

Our goal is to use machine learning to:

1. Predict hospital resource usage
2. Identify high-demand hospitals
3. Group hospitals based on efficiency

The goal is to help healthcare administrators make smarter, data-driven decisions.

Dataset Overview

1. We used a publicly available dataset on California hospitals.
2. The dataset contains over 578,000 records.
3. Key features we used in our project include:
   - Licensed, available, and staffed bed counts (BED_LIC, BED_AVL, BED_STF)
   - Number of days per report (DAY_PER)
   - Hospital name and county (FAC_NAME, COUNTY)
   - Type of service (Variable)
   - Total resource usage (Amount)
4. We focused mainly on data related to hospital operations, regional context, and service utilization.

CRISP-DM Process

- **Business Understanding**: Optimize hospital performance and planning
- **Data Understanding**: Explored key operational metrics
- **Data Preparation**: Cleaned, encoded, and scaled features
- **Modeling**: Trained and tested three ML models
- **Evaluation**: Compared model performance using R^2, RMSE, Accuracy
- **Interpretation:** Used PCA to visualize hospital clusters and highlight efficiency patterns.

Machine Learning Approach

We used three types of machine learning models:

- Regression – To predict resource usage (Amount)
- Clustering – To group hospitals by efficiency levels
- Classification – To predict if a hospital will face high demand

Models

Regression Models

Target: Amount (Resource Use)

Models Used:
- Linear Regression
- Random Forest
- XGBoost

Best Result:
Random Forest
RMSE: 61.6 million
R^2: 0.916
Relative Error: 0.37%

Clustering Models
Goal: Group hospitals by efficiency

Algorithms:
- K-Means
- DBSCAN (sample only)
- Agglomerative (sample only)

Best Output:
K-Means with 3 clusters
Visualized using PCA

Classification Models
Goal: Predict high demand (binary label using median Amount)

Models Tested:
- Logistic Regression
- Gradient Boosting
- Random Forest

Best Model:
Random Forest
83.5% accuracy

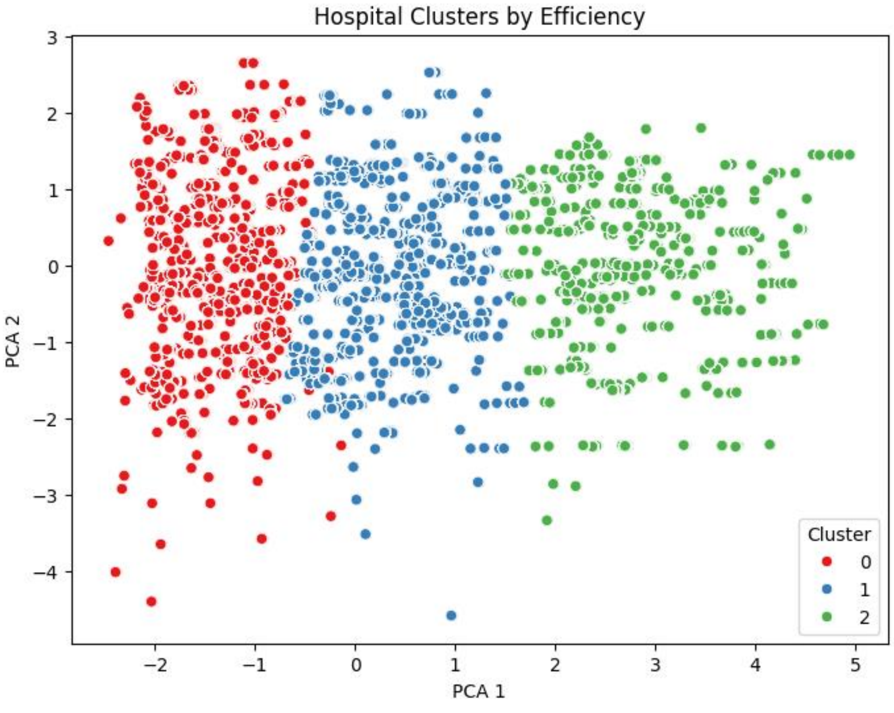
Regression Models

Algorithm	R^2 Score	RMSE (Millions)	Relative Error (%)	Notes
Linear Regression	0.0128	211.7	1.29%	Underfitting; poor predictive power
Random Forest Regressor	0.9164	61.6	**0.37%**	Best model for predicting Amount
XGBoost Regressor	0.8679	77.4	0.47%	Strong, but slightly behind RF

Clustering Models

Algorithm	Sample Size	Clusters Found	Notes
K-Means	Full	3	Clear clusters; interpretable PCA
DBSCAN	50,000	Mostly 1 + noise	Runtime constraints; not suitable for full dataset
Agglomerative Clustering	20,000	3	Good structure, but not scalable to full dataset

- **Clustering (K-Means) -** Visualized using PCA
- 3 Efficiency Groups:
  - Cluster 1: High capacity
  - Cluster 2: Moderate
  - Cluster 0: Underutilized



Classification Models (Without Outliers)

Algorithm	Accuracy (%)	Precision	Recall	F1-Score	Notes
Logistic Regression	54.2	0.54	0.54	0.54	Weak linear model baseline
Random Forest Classifier	60.3	0.61	0.60	**0.60**	Best without outliers
Gradient Boosting	58.0	0.58	0.58	0.58	Balanced but slightly lower than RF

Classification Models (With Outliers)

Algorithm	Accuracy (%)	Precision	Recall	F1-Score	Notes
Logistic Regression	54.2	0.54	0.54	0.54	No performance improvement
Random Forest Classifier	**83.5**	0.84	0.84	**0.84**	Excellent performance w/ outliers
Gradient Boosting	82.3	0.83	0.83	0.83	Strong alternative, close to RF

Limitations & Future Work

Limitations:

- DBSCAN, Agglomerative Clustering not scalable to full dataset

Future Work:

- Add time-based forecasting
- Build dashboard (Tableau/Power BI) for hospital insights

Final Thoughts

- ML models forecast hospital needs effectively
- Classification helps in early planning
- Clustering reveals operational disparities
- Our approach aligns with CRISP-DM and project goals

Conclusion

- Models can reliably forecast hospital needs and reveal performance gaps.
- Insights support better staffing, budgeting, and hospital policy decisions.

THANK YOU