PREDICTING HOSPITAL READMISSIONS TO OPTIMIZE HEALTHCARE RESOURCE ALLOCATION

Presentation by:

Shridhar Kumar
Kritika Gahlawat
Biswajit Gorai
Saswata Ghosh
Neha Rana

DIABETIC PATIENT READMISSION IS A SIGNIFICANT CHALLENGE IN HEALTHCARE MANAGEMENT. EARLY IDENTIFICATION OF HIGH-RISK PATIENTS AND OPTIMAL RESOURCE ALLOCATION ARE CRUCIAL FOR IMPROVING PATIENT OUTCOMES AND REDUCING HEALTHCARE COSTS.

Objectives

- Develop an interactive dashboard for diabetic readmission analysis
- Implement predictive models for risk assessment
- Create visualizations for data analysis and model performance
- Generate recommendations for resource optimization
- Provide actionable insights for healthcare professionals

Your paragraph text

DATASET OVERVIEW

- Dataset Size: 101,766 patient encounters
- Features: 50 columns covering demographic data, medical history, medications, diagnoses, and readmission info

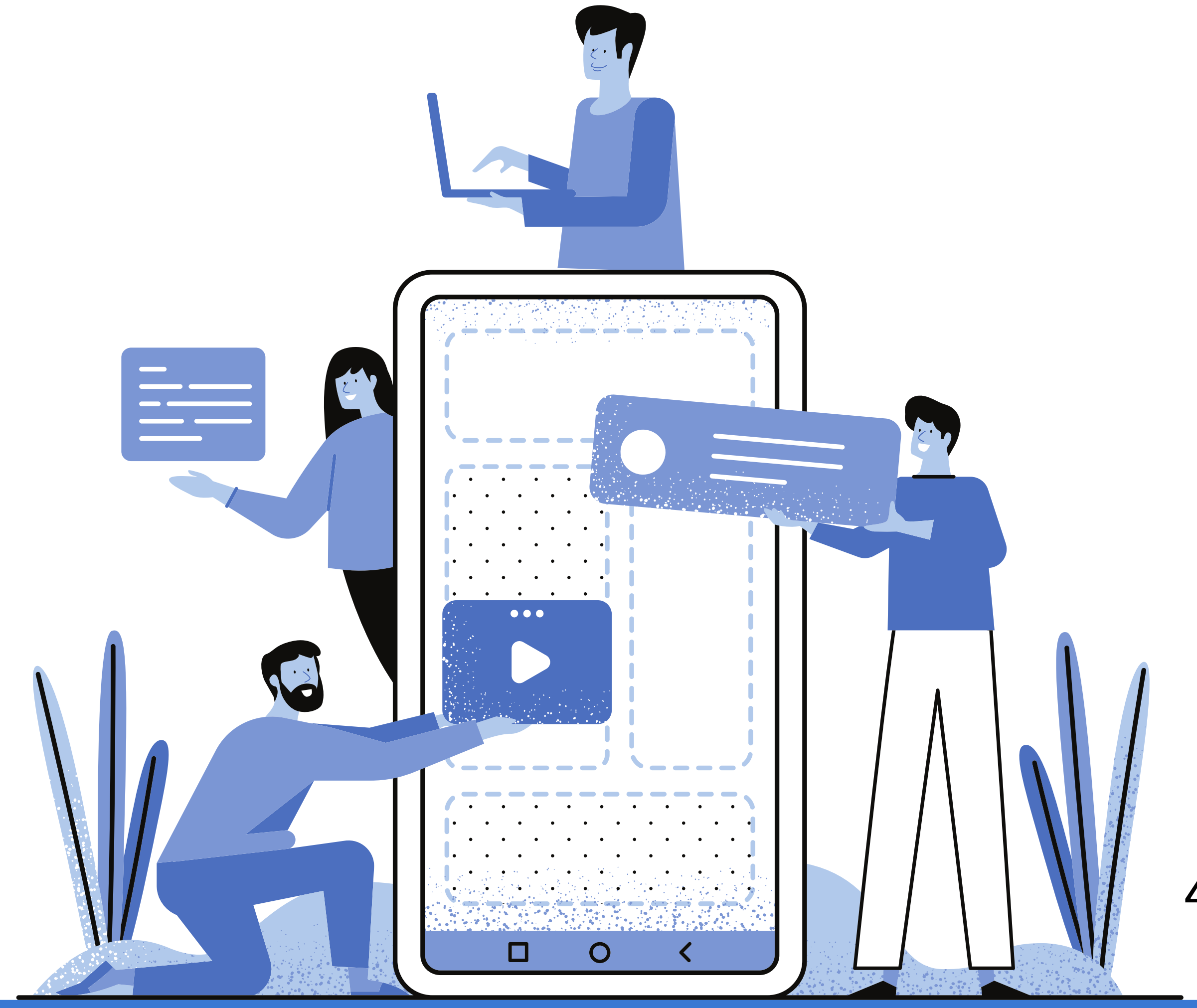
EXPLORATORY DATA ANALYSIS

Data overview

- Patient statistics
- Demographic analysis
- Medical feature distribution

Visualizations

- Readmission distribution charts
- Correlation heatmaps
- Feature importance plots



DATA PREPROCESSING AND FEATURE ENGINEERING

Absent data notion: Time ordering

	A	B	C	D	E
1	**patient id**	**encounter id**	**patient features**	**encounter features**	**readmitted**
2	A	random #	(data)	(data)	YES
3	A	random #	(data)	(data)	YES
4	A	random #	(data)	(data)	YES
5	A	random #	(data)	(data)	NO
6	B	random #	(data)	(data)	YES
7	B	random #	(data)	(data)	YES
8	B	random #	(data)	(data)	YES
9					

FINAL FEATURE SUMMARY

- Started with **47** features
- After dropping weakly or uninformative features and adding custom ones, there were 183 features, a mix of numeric and categorical
- After OneHotEncoder() transforms categorical features, a final set of **281** features is used for modeling

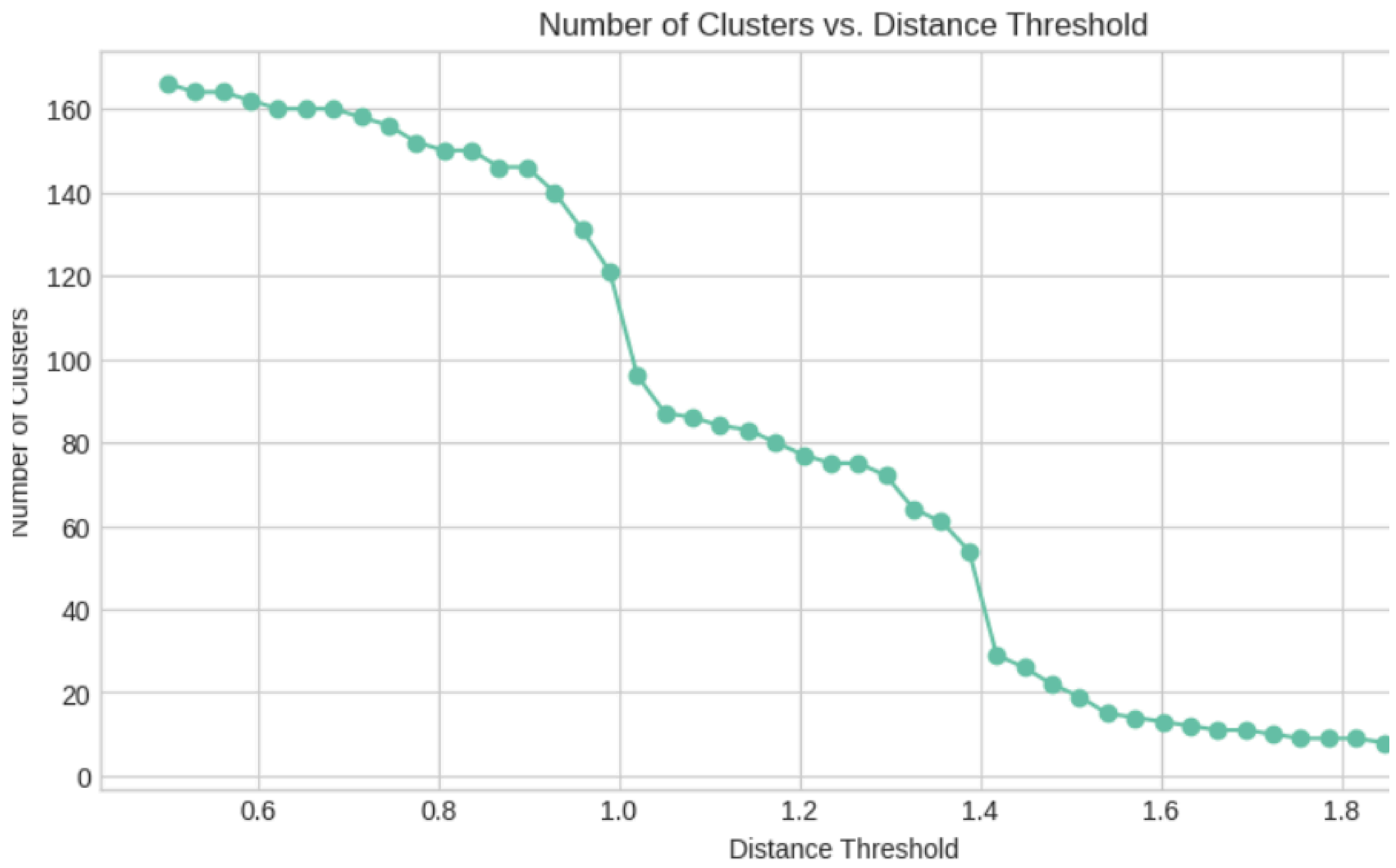
FINAL DATAPOINTS

RECORDS TRIMMED from model training:

- Patients expiring, or discharged to hospice (not expected to readmit)
- Patients under 20 years old, or over 80
- Patients with more than 13 admits in the 10-year period of data collection
- 101,766 - 23,023 = 78,743 (final N)

FEATURE SELECTION

Hierarchical Clustering



FINAL DATAPOINTS

After feature selection(Hierarchical Clustering method) we select 30 features

MODEL TRAINING

- Gradient Boosting Model
- Random Forest Model
- Logistic Regression Model
- Xgboost
- Light BGM
- Catboost
- Stacking

MODEL EVALUATION

- Accuracy
- Precision
- Recall
- F1-score
- ROC AUC

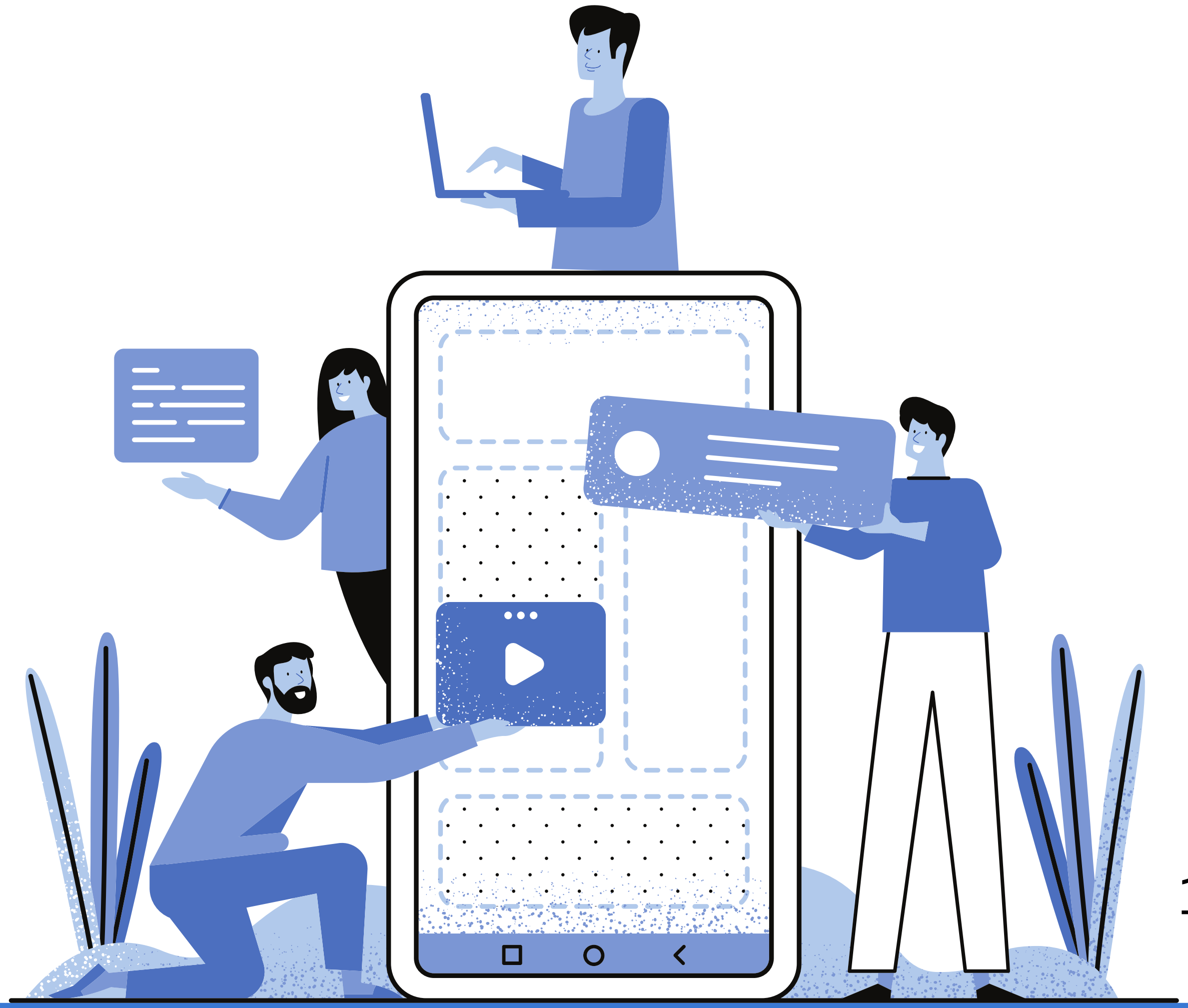
MODEL PERFORMANCE

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	0.8340	0.8511	0.7719	0.8095	0.9320
Random Forest	0.8883	0.8057	0.9959	0.8908	0.9330
Gradient Boosting	0.8878	0.8070	0.9929	0.8903	0.9340
XGBoost	0.8884	0.8065	0.9946	0.8910	0.9340
CatBoost	0.8881	0.8065	0.9933	0.8902	0.9330
LightGBM	0.8879	0.8072	0.9904	0.8894	0.9340
Stacking	0.8865	0.9686	0.8170	0.8864	0.9330
Voting	0.8868	0.9861	0.8084	0.8884	0.9330

PREDICTION

- Readmission risk prediction
- Predict risk analysis



RECOMMENDATION

<u>Recomendation about the resource allocation area</u>

- High Risk
- Medium Risk
- Low Risk

CONCLUSION

The Readmission Analysis Dashboard effectively tackles patient readmission and resource planning. Its intuitive design and rich features offer actionable insights for enhancing patient care. This project showcases the impactful use of data science in healthcare through a scalable, collaborative solution.

REFERENCES

1.Dash Documentation: https://dash.plotly.com/
2. Pandas Documentation: https://pandas.pydata.org/
3. Plotly Documentation: https://plotly.com/python/
4. Healthcare Analytics Research Papers
5. Machine Learning in Healthcare Resource

THANK YOU

Your paragraph text