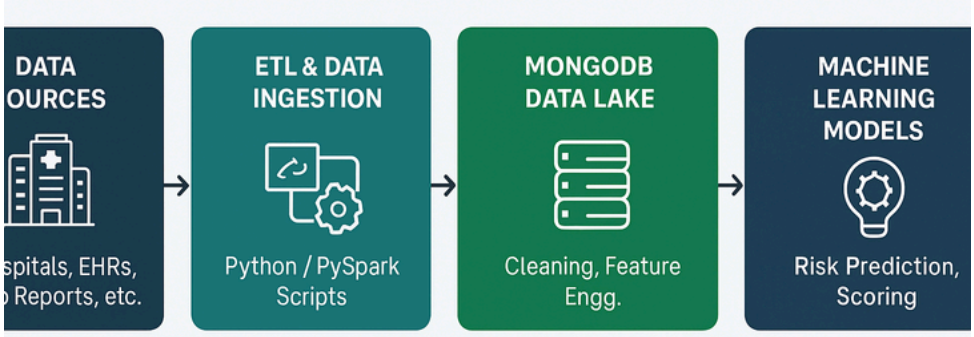
Predictive patient risk data lake

objective: To build a MongoDB-based Healthcare Data Lake that integrates patient history, appointments, and diagnosis data to predict high-risk patients using ML models enabling preventive care and data-driven decision making in hospitals.

Transforming data into actionable insights for early risk detection.

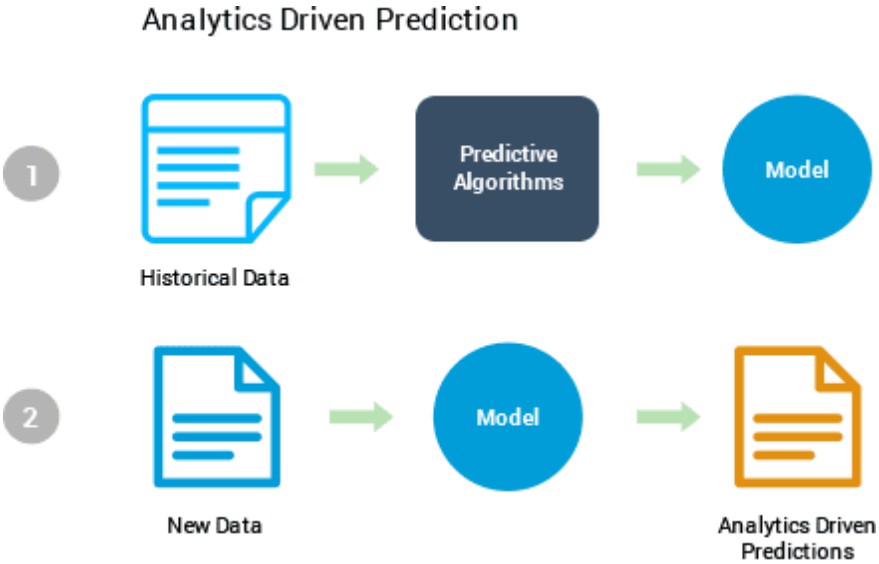
Features :
Unified data storage for all patient records
Scalable & schema-flexible MongoDB collections
Predictive insights for early diagnosis and intervention
Integration-ready with hospital management systems

Tech Stack:
Database: MongoDB Atlas
Processing: Python, PySpark
ML Models: Logistic Regression, Random Forest, XGBoost
Visualization: Power BI / Streamlit



Benefits & Industry Impact

- Reduces hospital readmissions
- Faster diagnosis through data-driven insights
- Supports personalized healthcare
- Improves hospital resource management
- Helps identify at-risk populations early



Key Features

- Scalable data storage for patient history
- ML-powered high-risk patient prediction
- Integration with real-time hospital systems
- NoSQL flexibility for diverse data formats
- Dashboard for actionable clinical decisions