

From Code to Champions: Our Journey to Winning the RWEsearch & Health AI Innovation Hackathon!

We are absolutely ecstatic to announce that our team, **PREDICTORS**, has secured **1st place** in the RWEsearch and Health AI Innovation Hackathon, organized by **Healthark**! This journey has been a whirlwind of data, code, and collaboration, and we are incredibly proud to see our hard work recognized.

Our team, composed of **Urva Gandhi**, **Harsh Khanna**, **Himanshu Mishra**, **Krish Marakana**, and **Hetvi Vanzara**, came together with a shared passion for leveraging technology to solve real-world problems. This hackathon provided the perfect arena to do just that.

Here's a look at our journey, from understanding the complex problem to building a winning solution.

The Challenge: Predicting the Future of Patient Care

The problem statement was both challenging and deeply impactful: design an AI-driven solution using anonymized data from the Centers for Medicare & Medicaid Services (CMS) to predict healthcare outcomes². The goal was to focus on patients with chronic conditions and provide early insights into factors like:

- Disease progression
- Hospital readmission rates
- Treatment costs and efficacy

Essentially, Healthark challenged us to move from reactive healthcare to a proactive, predictive model that could improve patient care and system efficiency.

Our Solution: An AI-Powered Crystal Ball for Healthcare

Our motivation was clear: chronic diseases account for over 75% of healthcare costs, and

frequent hospital readmissions are a sign of poor patient outcomes and a huge financial burden. We aimed to build a tool that could empower healthcare providers with scalable, actionable clinical insights.

Our project's primary objectives were to:

- **Predict hospital readmissions** within 30/60/90 day windows.
- **Track the progression of chronic diseases** like Congestive Heart Failure (CHF), Diabetes, and COPD.
- **Estimate treatment costs** and identify potential savings.

We developed an end-to-end predictive analytics platform, presented through an interactive and user-friendly dashboard.

A Quick Tour of Our Project

Our final output was an interactive web app designed for healthcare providers, analysts, and policymakers. Here's a glimpse of its core features:

1. **Data Overview & Dashboard:** The main dashboard provides a high-level view of the patient population, including the number of unique patients, inpatient admissions, and outpatient claims. It also displays key statistics on actual vs. predicted readmission rates for 30, 60, and 90-day periods, offering an immediate sense of the challenge at hand.
2. **Model Performance Visualized:** Transparency is key in AI. Our platform visualizes the performance of different machine learning models using ROC curves. This allowed us to demonstrate that our chosen model (XGBoost) consistently outperformed others like Logistic Regression and Random Forest in predicting readmissions, with an AUC score of up to 0.844 for 90-day readmissions.
3. **Feature Importance Insights:** The "why" behind a prediction is as important as the prediction itself. Our dashboard includes visualizations of "Random Forest Feature Importances," which clearly show the top factors driving readmission risk. Factors like the number of prior admissions, total inpatient cost, and specific diagnosis codes consistently ranked as top predictors.
4. **Chronic Condition Analysis:** We included a clear bar chart showing the prevalence of various chronic conditions within the patient population¹⁷. This helps stakeholders

quickly understand the landscape of patient health, identifying that conditions like Ischemic Heart Disease, Diabetes, and CHF are among the most common¹⁸.

Under the Hood: The Tech Stack

To bring this solution to life, we utilized a robust stack of modern data science and web development tools:

- **Backend & Machine Learning:** Python, Pandas, Scikit-learn, XGBoost.
- **Models Explored:** Logistic Regression, Random Forest, Gradient Boosting, and XGBoost.
- **Dashboard & Frontend:** Streamlit, for its power in rapidly creating interactive data applications.
- **Data Visualization:** Matplotlib and Seaborn for generating insightful plots and charts.

Our Blueprint: How We Solved the Problem

Our approach was systematic, mirroring the key tasks outlined in the hackathon problem statement:

1. **Data Analysis and Feature Engineering:** We started by diving deep into the CMS data, which included beneficiary details, inpatient/outpatient claims, and prescription events. We cleaned the data, handled inconsistencies, and engineered crucial new features like

number_of_hospitalizations, medication_adherence_score, and total_cost_per_patient.
2. **Model Development:** We adopted a multi-model strategy, training several algorithms from simple regressions to complex boosting trees²². We used a temporal data split (training on 2008-2009 data and testing on 2010) to simulate a real-world scenario where a model predicts future events²³.
3. **Outcome Prediction:** Our final models delivered high-accuracy predictions for 30, 60, and 90-day readmission risks. The output was a high-risk patient list that enables providers to intervene early.
4. **Insights and Visualization:** We focused on making our insights accessible. Instead of

just numbers, our dashboard tells a story, guiding users from a high-level overview to specific patient-level risks.

5. **Actionable Recommendations:** Going beyond prediction, we provided a framework for actionable strategies. This included recommendations for early interventions like care coordination, medication reminders, and policy-level suggestions to optimize resource allocation.

What We Learned: More Than Just a Competition

This hackathon was an incredible learning experience. As college students, it gave us a taste of what it's like to work in a professional, agile environment.

- **Real-World Problem Solving:** Moving beyond textbook examples to work with messy, complex CMS data was a challenge that taught us the importance of robust preprocessing and feature engineering.
- **Teamwork and Collaboration:** We learned to effectively divide tasks, manage our time under pressure, and merge our individual skills to build a coherent, powerful solution. Our success is a testament to our teamwork.
- **The Power of a Plan:** From gathering requirements from the problem statement to designing our architecture and planning our presentation, we realized that a solid plan is the foundation for any successful project.
- **Making an Impact:** Most importantly, we worked on a problem that matters. The solution we built has the potential to genuinely improve patient outcomes and support the healthcare system.

We are immensely grateful to Healthark for this opportunity and are looking forward to celebrating this victory in Ahmedabad! This win has not only boosted our confidence but has also ignited our passion to continue building innovative solutions for a healthier future.