Hackathon Use Cases – Life Sciences

RMK Use Cases for Life Sciences

S.no	Use Case	Data Source	Use case Description / Problem Statement	Objective	Type
1	Disease Detection – Tumors	https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/data	Early detection and classification of brain tumors is an important research domain in the field of medical imaging and accordingly helps in selecting the most convenient treatment method to save patients life. When benign or malignant tumors grow, they can cause the pressure inside your skull to increase. This can cause brain damage, and it can be life-threatening.	Early detection of Tumor and it's class from MRI images by use of advance AI techniques like image processing and deep learning.	Image processing / Computer vision
2	Predicting Hospital Readmissions	https://www.kaggle.com/datasets/dubradave/hospital-readmissions	Hospital readmissions are one of the costliest challenges facing healthcare systems, but conventional models fail to predict readmissions well. Many existing models use exclusively manually-engineered features, which are dataset-specific.	Build a predictive model that can accurately classify patients as readmitted or not, based on the provided features. This model can help healthcare providers identify high-risk patients and take proactive measures to prevent readmissions.	Classification
3	Predicting medical equipment failure	https://www.kaggle.com/datasets/vladimirmijatovic/faulty-medical-devices-global-dataset	Predicting medical device failure is crucial for ensuring patient safety, minimizing downtime, and reducing maintenance costs. This use case involves leveraging data analytics and machine learning to predict potential failures of medical devices before they occur.	Develop a machine learning algorithm that can predict medical device failures, enabling proactive interventions and maintenance. The data consist of large collection of recalls, safety alerts and field safety notices about medical devices distributed worldwide.	Classification
4	Claims Fraud Detection Analysis	https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis	Provider fraud is one of the biggest problems being faced by Medicare program. According to the government, the total Medicare spending increased exponentially due to frauds in Medicare claims. Healthcare fraud is an organized crime which involves peers of providers, physicians, beneficiaries acting together to make fraud claims.	The goal is to predict the potentially fraudulent providers based on the claims filed by them. Alongside, also discover important attributes helpful in detecting the behavior of potentially fraud providers. The dataset contains Inpatient claims, Outpatient claims and Beneficiary details of each provider.	Classification

RMK Use Cases for Life Sciences Contd…

S.no	Use Case	Data Source	Use case Description / Problem Statement	Objective	Type
5	Age Prediction	https://www.kaggle.com/datasets/mariafrenti/age-prediction/code	Accurately predicting an individual's age based on various data inputs is a challenging task due to the inherent variability and complexity of the data. Traditional methods of age verification can be intrusive, time-consuming, and prone to errors. Inadequate age prediction can lead to non-compliance with regulations, suboptimal personalization, and security vulnerabilities. Age prediction involves estimating an individual's age using various data inputs, such as images, text, voice, or other biometric information.	Using relevant machine learning techniques, train a model to predict a person's age with a low margin of error. Identify features and preprocessing that can be done to enhance the model. This capability can be utilized in a range of applications, from enhancing user experience to ensuring compliance with age-related regulations.	Computer vision
6	Sales Analysis & Forecasting	https://www.kaggle.com/code/milanzdravkovic/pharma-sales-data-analysis-and-forecasting/input	Pharma sales analysis and forecasting involve the use of data analytics to understand past sales performance, identify trends, and predict future sales. This helps pharmaceutical companies optimize their sales strategies, inventory management, and resource allocation. Time series forecasting involves predicting unknown values based on historical data patterns. It's particularly useful for scenarios where data is collected over time, such as sales.	The objective is to validate different methods and approaches related to sales time series data preparation, analysis and forecasting, with aim to facilitate recommending sales and marketing strategies based on trend/seasonality effects and forecasting sales with diverse characteristics, such as stationarity, seasonality and sales data variance	Time series Analysis

RMK Use Cases for Life Sciences Contd…

S.no	Use Case	Data Source	Use case Description / Problem Statement	Objective	Type
7	Q&A Chatbot for documentation	PDF documents (Financial reports, brand.com etc.) https://www.rxabbvie.com/pdf/rinvoq_pi.pdf	In the healthcare domain, there is a growing need for accurate and up-to-date drug information. Healthcare providers, patients, and caregivers often seek detailed prescribing information, safety alerts, and patient-specific advice related to medications. To address this need, a GENAI enabled chatbot that can extract prescribing and patient information from drug websites and presents it in a user-friendly format can save lot of time for caregivers.	Build a drug information Chatbot that can assist its users by extracting prescribing information (dosage, administration, contraindications, etc.) from drug labels, Retrieving patient-specific advice (e.g., interactions, precautions) and generating user-friendly responses. Team can leverage the PDFs from brand websites (Humira.com, Skyrizi.com, etc.) and generate the response from chatbot based on user's query. Expected Chatbot Features 1. Extract any kind of PDF file 2. It should be contextually aware from previous interactions in the same session while responding subsequent queries 3. It should provide citations from PDF pages while responding	Gen AI
8	Semantic Search (PubMed)	API Details - https://www.ncbi.nlm.nih.gov/home/develop/api/	PubMed is a widely used online database maintained by the National Center for Biotechnology Information (NCBI). It provides access to a vast collection of biomedical literature, including research articles, clinical studies, reviews, and more. Researchers, clinicians, and students rely on PubMed to find relevant scientific information across various disciplines. Challenges with Searches in PubMed: 1. Keyword Overload 2. Boolean Operators Many users are unaware of Boolean operators 3. Truncation 4. MeSH Terms: Medical Subject Headings (MeSH) are standardized terms used to index PubMed articles.	Given a specific query (e.g., an article title, keywords, or a combination), design an approach to retrieve relevant PubMed records that match the query semantically. The goal is to enhance search precision by considering semantic relationships between terms rather than relying solely on exact keyword matches. Approach: Utilize the PubMed API (specifically, the ESearch API) to perform the search. Specify the search query using the appropriate parameters, such as the title field for exact title searches. Showcase the search results in Frontend	Gen AI

Hackathon Use Cases – Healthcare

RMK Use Cases for Healthcare

Use Case 1: Claims Fraud Detection and Prevention

Objective: Develop an AI system that identifies potentially fraudulent Medicare claims in real-time, reducing financial losses while maintaining care access for legitimate patients.

Domain Challenge: Healthcare fraud costs US payers $68-230 billion annually, with traditional rule-based systems generating 95%+ false positives.

Problem Statement: Build an intelligent system that analyzes claims patterns and provider behaviors to detect fraud with 90%+ accuracy while reducing false positives by 60%.

Datasets:

- **Primary:** CMS Synthetic Public Use Files (DE-SynPUF) - https://www.cms.gov/data-research/statistics-trends-and-reports/medicare-claims-synthetic-public-use-files/cms-2008-2010-data-entrepreneurs-synthetic-public-use-file-de-synpuf Provider Specific Data for Public Use in Text Format | CMS
- **Secondary:** Healthcare Provider Fraud Detection (Kaggle) - https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis
- **Tertiary:** CMS Provider Data - https://data.cms.gov/provider-data/ 2025 Marketplace Open Enrollment Period Public Use Files | CMS

Technical Requirements:

- Real-time scoring (<3 seconds per claim)
- Explainable AI for audit compliance
- ROI tracking (target: $10 saved per $1 invested)
- Support for Medicare Parts A & B claims

RMK Use Cases for Healthcare Contd…

Use Case 2: Member Risk Stratification and Care Management

Objective: Identify high-risk members early and deploy targeted interventions to prevent costly complications and hospitalizations.

Domain Challenge: 5% of Medicare members account for 50% of total costs, making accurate risk prediction crucial for effective care management.

Problem Statement: Develop a risk stratification platform that predicts member health deterioration with 85%+ accuracy and recommends personalized interventions.

Datasets:

- **Primary:** CMS Synthetic Beneficiary Data - https://www.cms.gov/data-research/statistics-trends-and-reports/medicare-claims-synthetic-public-use-files/cms-2008-2010-data-entrepreneurs-synthetic-public-use-file-de-synpuf Provider Specific Data for Public Use in Text Format | CMS
- **Secondary:** Heart Disease UCI Dataset - https://archive.ics.uci.edu/dataset/45/heart+diseaseHeart Disease - UCI Machine Learning Repository
- **Tertiary:** Pima Indians Diabetes Dataset - https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-databasePima Indians Diabetes Database | Kaggle

Technical Requirements:

- Risk scoring for 30/60/90-day windows
- Integration with care management workflows
- Member stratification into 5 risk tiers
- Intervention ROI tracking

RMK Use Cases for Healthcare Contd…

Use Case 3: Hospital Readmission Prediction and Prevention

Objective: Predict and prevent 30-day hospital readmissions through risk scoring and targeted post-discharge interventions.

Domain Challenge: Hospital readmissions cost Medicare $26 billion annually, with CMS penalties affecting hospital reimbursements under HRRP.

Problem Statement: Develop a readmission prediction system that identifies high-risk patients at discharge with 80%+ accuracy and coordinates post-acute care.

Datasets:

- **Primary:** Hospital Readmissions Data - https://data.cms.gov/provider-data/ CMS DataCMS
- **Secondary:** MIMIC-III Demo Dataset - https://physionet.org/content/mimiciii-demo/1.4/ MIMIC-III Clinical Database v1.4
- **Tertiary:** Heart Disease Dataset - https://archive.ics.uci.edu/dataset/45/heart+disease Heart Disease - UCI Machine Learning Repository

Technical Requirements:

- Real-time readmission risk scoring at discharge
- Post-discharge care coordination workflows
- Integration with hospital EHR systems
- Cost savings validation (target: 15% reduction)

RMK Use Cases for Healthcare Contd…

Use Case 4: Prior Authorization Automation

Objective: Automate prior authorization decisions for medical procedures and prescriptions, reducing administrative burden while ensuring appropriate care.

Domain Challenge: Prior authorization delays affect 94% of physicians, causing treatment delays and $31 billion in administrative costs annually.

Problem Statement: Create an intelligent prior auth system that processes 90%+ of requests instantly with 95%+ accuracy and maintains full audit trails.

Datasets:

- **Primary:** CMS Provider Utilization Data - https://data.cms.gov/provider-summary-by-type-of-service/medicare-physician-other-practitioners 2025 Marketplace Open Enrollment Period Public Use Files | CMS
- **Secondary:** Medical Expenditure Panel Survey (MEPS) - https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp
- **Tertiary:** FDA Drug Database - https://dailymed.nlm.nih.gov/dailymed/

Technical Requirements:

- Sub-5 second decision processing
- Medical necessity rules engine
- Appeals prediction (target: 80% accuracy)
- Integration with EMR systems via HL7 FHIR

RMK Use Cases for Healthcare Contd…

Use Case 5: Pharmacy Benefit Management Optimization

Objective: Optimize formulary decisions and drug utilization to control pharmacy costs while ensuring appropriate medication access.

Domain Challenge: US prescription drug spending exceeds $350 billion annually, with 15-20% annual cost increases requiring sophisticated PBM management.

Problem Statement: Develop a PBM optimization platform that reduces pharmacy costs by 12% while maintaining 95%+ member satisfaction with drug access.

Datasets:

- **Primary:** CMS Part D Prescription Data - https://data.cms.gov/provider-summary-by-type-of-service/medicare-part-d-prescribers 2025 Marketplace Open Enrollment Period Public Use Files | CMS
- **Secondary:** FDA Orange Book - https://www.fda.gov/drugs/drug-approvals-and-databases/approved-drug-products-therapeutic-equivalence-evaluations-orange-book
- **Tertiary:** DrugBank Open Data - https://go.drugbank.com/releases/latest DrugBank Release Version 5.1.13 | DrugBank Online

Technical Requirements:

- Real-time formulary impact analysis
- Therapeutic equivalence optimization
- Drug utilization trend prediction

Cost-per-member-per-month tracking

RMK Use Cases for Healthcare Contd…

Use Case 6: Provider Network Optimization

Objective: Optimize provider networks to ensure adequate access while controlling costs and maintaining quality standards.

Domain Challenge: CMS requires specific provider-to-member ratios and geographic access standards while balancing network costs with member satisfaction.

Problem Statement: Build a network optimization tool that ensures 95%+ access compliance while reducing network costs by 8-12% and maintaining quality.

Datasets:

- **Primary:** CMS Provider Data Catalog - https://data.cms.gov/provider-data/ Overall hospital quality star rating | Provider Data Catalog
- **Secondary:** Medicare Provider Utilization Data - https://data.cms.gov/provider-summary-by-type-of-service/medicare-physician-other-practitioners 2025 Marketplace Open Enrollment Period Public Use Files | CMS
- **Tertiary:** Hospital Quality Star Ratings - https://www.nber.org/research/data/centers-medicare-medicaid-services-cms-hospital-compare-data Hospital Inpatient Quality Reporting Program | CMS

Technical Requirements:

- Geographic access analysis (30-minute drive time)
- Provider quality scoring algorithms
- Network adequacy compliance validation
- Cost impact modeling

RMK Use Cases for Healthcare Contd…

Use Case 7: Chronic Disease Management Platform

Objective: Build a comprehensive platform for diabetes/hypertension management that monitors patient vitals and prevents complications.

Domain Challenge: Chronic diseases account for 90% of healthcare costs, with diabetes alone costing $327 billion annually in the US.

Problem Statement: Create a disease management platform that reduces diabetes/hypertension complications by 25% and prevents 30% of emergency visits.

Datasets:

- **Primary:** Pima Indians Diabetes Dataset - https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-databasePima Indians Diabetes Database | Kaggle
- **Secondary:** Heart Disease UCI Dataset - https://archive.ics.uci.edu/dataset/45/heart+diseaseHeart Disease - UCI Machine Learning Repository
- **Tertiary:** CDC BRFSS Survey Data - https://www.cdc.gov/brfss/annual_data/annual_data.htm

Technical Requirements:

- Mobile app for patient data entry
- Predictive models for complication risk
- Care plan recommendation engine
- Integration with wearable devices (simulated)

Mandatory U.S. Healthcare Payer Technical Requirements

Regulatory Compliance

- HIPAA Security Rule: All PHI encrypted (AES-256 at rest, TLS 1.3 in transit)
- CMS Interoperability Rules: Patient access API compliance
- 21st Century Cures Act: Information blocking prevention
- Medicare Program Requirements: CMS regulation adherence

Data Standards & Integration

- HL7 FHIR R4: All clinical data exchanges
- X12 EDI Standards: Claims and administrative transactions
- ICD-10/CPT Coding: Medical coding accuracy (99.5%+)
- API Security: OAuth 2.0 + JWT tokens

Performance & Quality Standards

- Real-time Processing: <3 seconds for critical decisions
- Scalability: Support for 100K+ members minimum
- Availability: 99.9% uptime requirement
- Data Quality: 99.5% accuracy standards
- Audit Compliance: Complete activity logging