Diabetes 130-US Hospitals Readmission Analysis

Strategic Patient Risk Stratification & Readmission Predictive Modeling

A comprehensive health informatics analysis of diabetic patient readmission patterns using the UCI Diabetes 130-US Hospitals dataset. This project develops the **Vitality Complexity Index (VCI)** — a custom risk stratification algorithm to identify high-risk patients and reduce hospital readmission rates.

Project Overview

This analysis examines **100,000+ patient encounters** spanning 130 US hospitals over 10 years (1999-2008) to identify key drivers of hospital readmission among diabetic patients. The project was developed for Vitality Health Network (VHN) to address challenges under the CMS Hospital Readmissions Reduction Program (HRRP).

Key Findings

- **46.9% Combined Readmission Rate** (11.3% within 30 days, 35.5% after 30 days)
- **8% Higher Readmission Risk** for insulin-dependent patients vs. oral medication users
- **4.3% Increased Risk** when medications are changed during hospital stay
- **60%+ of High-Risk Readmissions** originate from Emergency Department admissions

---

Repository Structure

```
├── VHN_Analysis.ipynb              # Main Jupyter notebook with full analysis
├── vhn_analysis_pipeline.py        # Python script version of the pipeline
├── VHN_Strategic_Insight_Report.md # Executive report for stakeholders
├── data_files/
│   └── data_files/
│       ├── diabetic_data.csv       # Main dataset (101,766 records)
│       └── IDs_mapping.csv         # ID-to-description mappings
├── plots/                          # Generated visualizations
│   ├── readmission_distribution.png
│   ├── age_distribution.png
│   ├── medication_efficacy_group.png
│   ├── vci_validation.png
│   └── ...
└── README.md
```

Analysis Pipeline

Phase 1: Data Sanitation

- Missing value analysis (96.8% weight data missing)
- Deceased patient removal (1,652 patients excluded for methodological rigor)
- Categorical re-engineering for clinical IDs

Phase 2: Web Scraping Enrichment

- Automated ICD-9 code lookup for top 20 diagnosis codes
- Integration of human-readable clinical descriptions

Phase 3: Exploratory Data Analysis

- Readmission distribution analysis
- Demographic profiling (age, race, gender intersectionality)
- Medication efficacy assessment
- Length of stay and discharge disposition analysis
- Correlation heatmap for multicollinearity validation

Phase 4: Feature Engineering - Vitality Complexity Index (VCI)

- Custom L.A.C.E-inspired scoring algorithm:
  - **L**: Length of Stay (0-7 points)
  - **A**: Admission Acuity (0-3 points)
  - **C**: Comorbidity Count (0-5 points)
  - **E**: Emergency Visit History (0-5 points)
- Risk stratification: Low (<7), Medium (7-10), High (>10)

Technologies Used

- **Python 3.x**
- **pandas** - Data manipulation and analysis
- **NumPy** - Numerical computing
- **Matplotlib & Seaborn** - Data visualization
- **Requests & BeautifulSoup** - Web scraping for ICD-9 enrichment
- **Jupyter Notebook** - Interactive analysis environment

Installation & Setup

1. Clone the repository:

```
git clone https://github.com/yourusername/diabetes-130-us-hospitals-analysis.git
cd diabetes-130-us-hospitals-analysis
```

2. Install dependencies:

```
pip install pandas numpy matplotlib seaborn requests beautifulsoup4
```

3. Run the Jupyter notebook:

```
jupyter notebook VHN_Analysis.ipynb
```

Dataset

Source: UCI Machine Learning Repository - Diabetes 130-US Hospitals

- **Records:** 101,766 patient encounters
- **Features:** 50 clinical and demographic variables
- **Time Period:** 1999-2008
- **Hospitals:** 130 US hospitals

Strategic Recommendations

1. **High-Risk VCI Outreach Protocol** — Mandatory 48-hour follow-up for patients with VCI >10
2. **EHR Integration** — Traffic light visualization (Red/Yellow/Green) on patient census boards
3. **Pharmacist-Led Medication Counseling** — Mandatory discharge education for medication changes

Projected Impact: $1.8M - $3.2M annual savings in avoided penalties and optimized bed-days

---

Author

Shehan Anujaya

License

Acknowledgments

- UCI Machine Learning Repository for the Diabetes 130-US Hospitals dataset